AD-786 878

DELPHI ASSESSMENT: EXPERT OPINION, FORECASTING, AND GROUP PROCESS

RAND CORPORATION

PREPARED FOR
DEPUTY CHIEF OF STAFF, RESEARCH AND DEVELOPMENT
(AIR FORCE)

APRIL 1974

R-1283-PR
April 1974

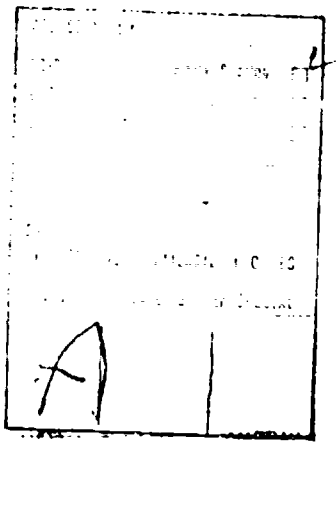
Delphi Assessment: Expert Opinion, Forecasting, and Group Process

H. Sackman

A Report prepared for

UNITED STATES AIR FORCE PROJECT RAND

Rand
SANTA MONICA, CA 90406



The research described in this Report was sponsored by the United States Air Force under Contract No. F44620-73-C-0011 — Monitored by the Directorate of Operational Requirements and Development Plans, Deputy Chief of Staff, Research and Development, Hq USAF. Reports of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

PREFACE

The Delphi technique originated at The Rand Corporation in the late 1940s as a systematic method for eliciting expert opinion on a variety of topics, including technological forecasting. Over the years, Rand has conducted a number of Delphi experiments. In addition, hundreds of Delphi studies have been published under corporate, government, and academic sponsorship, covering a vast range of topics, in the United States and abroad, including Europe, the Soviet Union, and Japan.

This report presents a critical analysis of the Delphi technique. The analysis is in four parts. First, the scope of the inquiry is defined, and issues pertinent to an evaluation of Delphi are raised. Second, conventional Delphi is evaluated against established professional standards for opinion questionnaires, and against associated scientific standards for experimentation with human subjects. Third, Delphi is evaluated with respect to its assumptions, principles, and methodology. Fourth, conclusions of the analysis are brought together and recommendations are made for the future use of Delphi.

A critical evaluation of Delphi is long overdue, and it is appropriate that such an analysis be conducted by Rand. The research undertaken for this report was supported by the U.S. Air Force under Project RAND. Although portions of the discussion involve technical points of sampling, psychometric experimental design, and statistical analysis, the analysis should be generally useful to Air Force and other planners who may be contemplating the use of Delphi.

CONTENTS

SUMMARY

This report presents a critical analysis and evaluation of the Delphi technique. To keep the analysis within manageable bounds, the prime focus is on Delphi method; the diverse application areas are secondary. The prologue defines the scope and organization of the inquiry and sketches key methodological issues associated with the complete cycle of conventional or characteristic Delphi studies. The discussion proceeds to an evaluation of Delphi against professional standards for social experimentation and for opinion questionnaires established by the American Psychological Association and other national professional organizations. Analysis of conventional Delphi indicates that it does not satisfactorily meet the numerous experimental and methodological standards cited for test design, item analysis, subject sampling, reliability, validity, administration, interpretation of findings, and warranted social use.

The main body of the critique reviews methodological principles and key assumptions associated with Delphi. This analysis reveals: considerable evidence that results based on the opinions of laymen and "experts" are indistinguishable in many cases; aggregate raw opinion presented as systematic prediction; technical shortcomings, such as untested and uncontrolled halo effects in the application of Delphi questionnaires; unsystematic and non-replicable definition and use of "experts;" manipulated group suggestion rather than real consensus; ambiguity in results stemming from vague questions; acceptance of snap judgments on complex issues; and the virtual absence of a vigorous critical methodological literature even though hundreds of Delphi studies have been published. The accuracy of the technique, in generating forecasts and other "expert" estimates, is necessarily suspect so long as Delphi questions are not empirically linked to objective and independently verifiable external validation criteria. These liabilities are counter-balanced primarily by a popular demand for systematic expert opinion, and by the convenience, low cost, and simplicity of the method. It is argued that such advantages are inconsequential if the Delphi concept, method, and results are inherently untrustworthy.

The analysis concludes that conventional Delphi is basically an unreliable and scientifically unvalidated technique in principle and probably in practice. In the absence of a comprehensive survey of the extensive applications literature, it is suggested, but not proven, that the results of most Delphi experiments are probably unreliable and invalid. Even variations of conventional Delphi should not be encouraged unless they explicitly attempt to meet the challenge of generally accepted standards of rigorous empirical experimentation in the social sciences. Except for its possible value as an informal exercise for heuristic purposes, Delphi should be replaced by demonstrably superior, scientifically rigorous questionnaire techniques and associated experimental procedures using human subjects.

As the preferred alternative to conventional Delphi, professionals, funding agencies, and users are urged to work with psychometrically trained social scientists who can apply rigorous questionnaire techniques and scientific human experimentation procedures tailored to their specific needs. The final recommendation is that conventional Delphi be dropped from institutional, corporate, and government use until its principles, methods, and fundamental applications can be established experimentally as scientifically tenable.

ACKNOWLEDGMENTS

This project was instigated by Barry Boehm, who was firmly convinced that a Rand-supported evaluation of Delphi was long overdue. The author is indebted to Allan Haile for numerous discussions in planning and executing this project, and for encouraging the development of some key ideas in this critique. Barbara Quint provided invaluable library services and helped prepare the semi-annotated Delphi literature compilation listed in the Appendix. The author has had the benefit of extensive informal Rand feedback with detailed technical reviews of the draft manuscript, including two lively "adversary" seminars.

The author is particularly indebted to Ned Feldman, who volunteered the constructive criticism that formed much of the basis for the extensive overhaul and expansion of the initial draft. The timely editorial services of Brownlee Haydon helped smooth many of the rough edges throughout the manuscript, for which the author is most grateful. Peter Weiner, who inherited the responsibility for overseeing the final disposition of this investigation, insisted upon a more balanced approach to difficult, controversial material throughout this report; the orientation of the final manuscript reflects this influence.

Because the author has not been able to find seriously critical literature of any depth on Delphi, he has had to assume the full burden of the explication and defense of his critique over a wide-ranging spectrum of Delphi issues. The final evaluation of Delphi in this report is strictly the responsibility of the author, who was free to develop his own position, unfettered by constraints.

1. PROLOGUE

This report does not attempt to review the history of the Delphi method; its primary purpose is to present a critique of Delphi. A general description of the essential features of Delphi is presented in the following section.

Organizing a meaningful critique of Delphi presented many problems. After considering various alternatives, a four-step schema was adopted, corresponding to the organization of this paper. First, raise the various types of definitive issues pertinent to a Delphi critique. Second, evaluate conventional Delphi against established professional standards for opinion questionnaires and scientific standards for empirical experimentation with human subjects. Third, evaluate Delphi in terms of its unique assumptions, principles, and methodology. Finally, summarize basic conclusions and make recommendations concerning the future use of Delphi.

Before proceeding with the critique, three caveats on the scope and limitations of this study should be made explicit. Delphi has been used for a vast array of applications in business, science, education, medicine, and other areas, both broad and specialized. The total literature has been estimated to include several hundred titles; a substantial number of these are proprietary or otherwise inaccessible. The author has been able to examine approximately 150 Delphi studies conducted at Rand and elsewhere (see the semi-annotated listing of Delphi and related publications in the Appendix). The author makes no claim to having examined all the literature, particularly all the applications literature.

The focus of this study is on Delphi principles and methodology. The literature that has been reviewed contains the basic writings of the originators and key practitioners of Delphi, both within and outside Rand, with critical coverage of Delphi principles, assumptions, and procedures. Evaluative inferences from methodology to application are admittedly based on illustrative examples rather than on direct examination of all relevant studies. The validity of such

inferences should be judged on the coherence of arguments put forth and the representativeness of examples used.

Another constraint is the elusiveness of a fixed, universally agreed upon, working definition of Delphi. Many variants have emerged, some departing widely from the Delphi procedure associated with its Rand origins. An attempt is made in the next section to present a definition and characterization of "conventional Delphi." The use of the term "Delphi" in this report refers primarily to "conventional Delphi," which may or may not apply to Delphi variants, depending upon the issues and the context.

A third caveat on the scope of this study is that it does not compare Delphi systematically with competing techniques. A comparison of Delphi with such techniques as simulation, trend extrapolation, gaming, morphological models, scenarios, relevance trees, input-output tables, contextual mapping, brainstorming, dialectical planning, critical path methodology, etc., would require an independent review and evaluation of each of these techniques and the systematic comparison of each with the others for key objectives and application areas. Undoubtedly, such a comprehensive critical appraisal of the methodology of the entire field of forecasting and planning techniques is long overdue, for much the same reasons that an in-depth Delphi critique is overdue. (For an instructive initial comparison and rating of these and related techniques, and for an appreciation of the magnitude of the task, see Rosove, 1967, and Sackman and Citrenbaum, 1972). However, such an appraisal would involve an effort an order of magnitude larger than the present project. As desirable as such an undertaking might be, this evaluation is necessarily limited to a comparison of conventional Delphi with scientific questionnaire development and experimental methodology with human subjects, and to questioning many of the basic assumptions and methods of the technique as it is currently being applied.

2. DELPHI METHOD: ISSUES, PROBLEMS, AND DEFINITION

The objective in this section is to identify certain methodological issues, and to characterize "conventional Delphi" for the purposes of this critique. The chronological framework for a Delphi study follows a problem-solving sequence: establishment of objectives, formulation of the problem, solution testing, and the writeup and dissemination of results. In the Delphi context, objectives include needs, goals, basic value assumptions, and expected payoffs. Formulation of the problem is accomplished through the design of the questionnaire and its experimental implementation. Solution testing includes iterative field administration and scoring of responses to the questionnaire. The last stage involves the interpretation of results by the Delphi director in communicating findings to others. Each stage is briefly examined to provide a chronological chain of methodological issues as a framework for this evaluation.

2.1. Delphi Objectives

Early Delphi studies at Rand were primarily concerned with scientific and technological forecasting. They were viewed as experiments with what was thought to be an interesting, and possibly useful, new technique. From these humble beginnings, Delphi has spread rapidly, with hundreds of studies appearing in the United States, accompanied by growing use in other countries, including extensive use in the United Kingdom (Currill, 1972), recent use in the Soviet Union (Martino, 1973), and in Japan. Delphi applications have grown in all directions to include forecasting of many social phenomena, including human attitudes and values (Reisman et al., 1969), and even the "quality of life" (Dalkey, Rourke, Lewis and Snyder, 1972). A large and growing roster of major firms have used Delphi for diverse purposes (see bibliographic Appendix). Applications have expanded until they are virtually indistinguishable from the questionnaire technique, broadly considered. Advocates, such as Turoff (1971), have expanded the scope of Delphi as a general-purpose vehicle for distributed human

communication and consensus, and for group problem-solving. Delphi has been propelled at an increasingly accelerated rate into the general field of questionnaire design and development not only for "experts," but for non-experts as well. The core question arises, how does Delphi rate in comparison with competing approaches in the well established fields of questionnaire design and application in the social sciences?

The payoff of a Delphi study is typically a presentation of observed expert concurrence in a given application area where none existed previously. This assumes that participating panelists are experts in the subject area, and that the reported consensus was obtained through reliable and valid procedures. Proponents of Delphi (Dalkey, 1969) stress three quintessential attributes that contribute to authentic consensus and valid results--anonymity of panelists, statistical response, and iterative polling with feedback. Is the trust placed in these central assumptions warranted?

In any decision to use delphi, there are various cost-effectiveness considerations. How much does a Delphi study cost in time and effort for the director and panelists, and how are such investments related to the usefulness of the final results? An associated issue is the attractiveness of Delphi as a quick and easy way to solicit rational expert opinion in an unknown area. Do such positive payoffs exist?

2.2 Formulation of the Problem

The next step in a Delphi study is the formulation of the problem, the design of the questionnaire and its application. How effectively is the area of inquiry defined and delirdted by the Delphi investigator? Is there an effort to make questionnaires bias-free? Are his assumptions spelled out? Are there explicit hypotheses and are they operationally defined? Has the relevant literature been reviewed and systematically evaluated? Have baseline statistics and qualitative characteristics of the area of inquiry been documented and spelled out, so that respondents derive their forecasts and opinions from a common specification of the current state-of-the-art?

In developing the questionnaire, many technical considerations arise. Is the questionnaire an informal, ad hoc collection of items,

or is it systematically designed as a standardized instrument to be administered under rigorously controlled conditions? How are the items constructed? How large was the original pool of items, how were they derived, and what pilot procedures were used for item analysis to prune them down to the final set used for the study? What psychometric scaling approach was selected (e.g., Thurstone, Likert, or Guttman psychometric scales, or econometric scales, see Pill, 1971) and what factors determine the selection?

Then there are problems concerning the panelist sample to which the questionnaire is applied. What is an "expert" in the target application field, and how are such experts operationally defined? How many panelists are used, and what are the expected levels of statistical precision of the results relative to planned sample size for the dispersion of responses anticipated? Can the selected panelist sample be systematically related to an objectively defined population with measurable sampling parameters? Is the choice of experts random, or is it selective? Are sampling procedures rigorously defined (see Cochran, 1963) relative to hypothesis testing for opinion polling?

2.3 Solution Testing

In administering the questionnaire, many problematic issues arise. How are dropouts handled in the results? Which items should be dropped, modified, or retained in their original form in successive Delphi rounds? What kind of feedback, how much feedback and in what form, should be presented to panelists? When is the point of diminishing returns reached in successive iterations? How long should the intervals be between successive rounds, and how can participants be encouraged to respond promptly to expedite turnaround time? What is the tradeoff between more items and a longer form versus fewer items with less data in relation to study objectives? Does the director reinforce and encourage conformist or dissenting behavior in successive rounds? In working with distributed Delphi by mailed questionnaires and iterative polling, what opportunities exist for misusing the technique?

2.4. Interpretation and Use of Results

In the final stage of writeup and dissemination of results, the main problems center around the analysis and interpretation of findings. Should only descriptive results be presented, or should all statistics be accompanied by standard errors of estimate, clearly indicating the empirical level of precision? Is it misleading to present only interquartile ranges in graphic portrayal of Delphi results, or should the full range and true dispersion of results also be presented? Should first-round results be presented showing the full dispersions of expert opinion? How strongly should the expert halo effect be exploited, or should it be controlled in evaluating results? Should the procedure and the interpretation give weight to adversary or consensus positions?

How strongly should procedural, administrative, statistical, and experimental limitations be stressed in the final publication? Are results put forth as scientific prediction or as conglomerate opinion? Has provision been made for replication testing or validity generalization in follow-on studies?

2.5. Definition of Conventional Delphi

The above review of the Delphi cycle provides a backdrop for the characterization of "conventional Delphi" as it is used in this critique. These are briefly described below under the categories of objectives, subjects, and techniques.

The application objective of conventional Delphi may be the forecasting of specified events, long-term or short-term; it may be the generation of quantitative estimates (e.g., costs, market demand, numbers of users, etc.) from a set of participants; or it may be aimed at qualitative evaluations (e.g., qualitative scales of agreement, disagreement, preferences among alternatives). The range of application objectives thus includes any type of quantitative or qualitative rating scale, and as such is coextensive with questionnaires broadly considered.

Other key objectives for conventional Delphi may be singled out, including consensus of participants and heuristic goals. The consensus intent of Delphi is typically oriented toward controlled and rational

exchange of iterated opinion leading toward optimal covergence of opinion achievable within the framework of the technique. The heuristic objective views Delphi as an educational technique to help participants, the director, and users to explore a problem area more thoroughly, leading to greater insight on the target problem.

Turning now to subjects, conventional Delphi is primarily concerned with experts, but may also use other subject groups who may be informed to a greater or lesser extent in the target area of inquiry, but who do not qualify as experts. Although this report focuses on the Delphi concept of expert, it is also directed at the growing use of non-experts. More broadly, this critique is concerned with the operational sampling procedures used in selecting Delphi subjects, expert or otherwise.

The technique category is the most detailed. Conventional Delphi, as used in this report, exhibits the following characteristics:

a. The format is typically, but not always, a paper-and-pencil questionnaire; it may be administered by mail, in a personal interview, or at an interactive, online computer console. The basic data presentation and data collection technique is the structured, formal questionnaire in each case.
b. The questionnaire consists of a series of items using similar or different scales, quantitative or qualitative, concerned with study objectives.
c. The questionnaire items may be generated by the director, participants, or both.
d. The questionnaire is accompanied by some set of instructions, guidelines, and ground rules.
e. The questionnaire is administered to the participants for two or more rounds; participants respond to scaled objective items; they may or may not respond to open-end verbal requests.
f. Each iteration is accompanied by some form of statistical feedback which usually involves a measure of central tendency, some measure of dispersion, or perhaps the entire frequency distribution of responses for each item.

g. Each iteration may or may not be accompanied by selected verbal feedback from some participants, with the types and amounts of feedback determined by the director.

h. Individual responses to items are kept anonymous for all iterations. However, the director may list participants by name and affiliation as part of the study.

i. Outliers (i.e., upper and lower quartile responses) may be asked by the director to provide written justification for their responses.

j. Iteration with the above types of feedback is continued until convergence of opinion or "consensus" reaches some point of diminishing returns, as determined by the director.

k. Participants do not meet or discuss issues face-to-face, and they may be geographically remote from one another.

It should be apparent that a one-sentence or even one-paragraph definition of "conventional Delphi" is not possible without leaving out many significant details and qualifications that receive substantial attention in this report. Generally speaking, the working definition of Delphi for this study embodies the "quintessential" model originating at Rand, with many related variations that more or less follow the iterative questionnaire format with anonymous statistical feedback.

This completes the review of issues raised by the conventional Delphi cycle and permits an evaluative comparison of Delphi with professional standards for opinion questionnaires and experimentation with human subjects.

3. DELPHI VERSUS SOCIAL SCIENCE STANDARDS

This section presents key standards in professional questionnaire design and use, and shows how Delphi measures up to them. The evaluative criteria are quoted from "Standards for Educational and Psychological Tests and Manuals," published by the American Psychological Association, (1966). This publication was jointly prepared by a committee representing three national organizations: The American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education. This committee worked over a period of five years in conjunction with numerous measurement specialists and test publishers.

The manual is currently undergoing revision under the auspices of the APA Office of Scientific Affairs. Public hearings on the proposed draft have been held in Washington, D.C. The provisional table of contents of the proposed version is shown in Table 1.

It cannot be too strongly emphasized that these guidelines represent responsible efforts to establish exemplary scientific standards in a controversial area with a history of continuing abuse on the part of some test developers, and with a history of continuing misunderstanding and under-education on the part of the public. Whether or not the reader identifies himself as a social scientist, he should be aware that there is a vast and highly germane literature reflecting an organized professional effort to serve the public interest.

Buros (1965), after dedicating a distinguished lifetime to professional quality control in the public domain for the testing field, concluded that only partial success is possible with the inevitable collusion between test promoters and a gullible public that expects far more from tests than they can possibly deliver. The carryover to Delphi, as this report shows, is more than mere coincidence. In the absence of any tradition with such guidelines, Delphi practitioners, participants, and users can neglect such standards only at their own peril.

Some may still argue that Delphi is not a conventional test,

Table 1

STANDARDS FOR DEVELOPMENT AND USE OF
EDUCATIONAL AND PSYCHOLOGICAL TESTS

TABLE OF CONTENTS

though it usually assumes the form of an iterative paper-and-pencil questionnaire. As such, they argue, Delphi is exempt and the APA Guidelines should not apply. A cursory review of the selected items, however, reveals that the guidelines deal with bedrock questions concerning sampling, controls, reliability of measures, and criterion validity which are universal to all scientific experimentation with human subjects. If Delphi is to be treated seriously as a professional technique, it must be judged by basic, minimum standards applicable to all empirical social science.

The historical precursors of Delphi in the opinion polling and social psychological literature were most explicit in applying rigorous questionnaire design and sampling techniques against the methods and findings of their studies. Cantril (1938) and McGregor (1938), in independent studies on predictions of social events, emphasized the severe limitations of questionnaire format and procedures and in the representativeness of subject sampling for any generalizations of their results. Kaplan, Skogstad, and Girshick, in a landmark study on "The Prediction of Social and Technological Events" (1950), presented a detailed listing of sampling, reliability, and validity problems encountered in this field in relation to rigorous questionnaire and polling standards. As a direct historical offshoot of these pioneering efforts, the Delphi technique does not possess or warrant any special dispensation exempting it from such scientific standards.

Delphi proponents may protest that concern with experimental method in the application of Delphi questionnaires is "misguided" because Delphi is a tool, and a tool, once developed, does not have to be experimentally administered each time it is used. This may be fine for weighing scales, rulers, compasses, spectrometers, voltmeters, and other measurement instruments frequently used in the physical sciences. However, it does not apply to questionnaires, or to paper-and-pencil testing broadly considered, or to Delphi in particular. A questionnaire is reliable and valid only to the extent that it is administered under conditions that replicate the basic experimental controls under which it was originally designed, tested, and validated. This means that each administration of the questionnaire is viewed as an

experimental replication of operationally designed conditions for individual response, collection of data, scoring, and interpretation. The layman's failure to realize that questionnaire tests are replicated experiments leads to abuses of the technique, non-comparability of results, and a general increase in measurement error variance.

Delphi iteration of questionnaires with feedback is a definitive empirical experimental procedure with human subjects in its own right. Neglect of standard experimental guidelines may lead to uncontrolled variations in results and inability to define, replicate, and validate methods and findings. This neglect may be acceptable for an informal exploratory technique, but it is unacceptable for a rigorous social science experiment. The compounding of methodological problems generated by an unscientific approach to the conduct of Delphi studies is described and illustrated in this section.

3.1. Scope of Standards

While the standards are quoted verbatim from the current manual, the author is fully responsible for the evaluative Delphi commentary. The manual covers paper-and-pencil testing broadly considered and, obviously, many of the standards do not pertain directly to Delphi. In what follows, a representative subset of key standards relevant to Delphi is cited, accompanied by evaluative commentary. The citations cover introductory, interpretive, validity, reliability, and administrative/scoring standards, taken from applicable sections in the APA manual.

In the direct quotes that follow, material is reproduced verbatim except for one term. The word "manual" is replaced by "test documentation." This is done because it was found that individuals unfamiliar with psychometrics found it difficult to understand the scope and intent of test "manuals." "Test documentation," which refers to test materials, instructions, controls, and reports of empirical results, norms, interpretations, and recommendations for use, is less likely to cause unintentional confusion for the layman in relating the guidelines to Delphi.

In the intoduction, the manual states:

> "These recommended standards cover not only tests as narrowly defined, but also most published devices for diagnosis, prognosis, and evaluation . . ."
>
> "The present standards apply to devices which are distributed for use as a basis for practical judgments rather than solely for research. Most tests which are made available for use in schools, clinics, and industry are of this practical nature" (p. 3).

Conventional Delphi studies, as applied prognostications, or as predictions, or as predictions of technological and social developments for a variety of end users, fall under the general purview of the manual.

3.2. Interpretive Standards

From the section on interpretation of findings, two items are selected. Ratings accompany each standard listed in the manual. Ratings are "essential," "very desirable," or "desirable," shown in caps at the end of each item.

> B4.2. When the statistical significance of a relationship is reported, the statistical report should be in a form that makes clear the sensitivity or power of the significance test. ESSENTIAL (p. 11).

Statistical significance is rarely reported in Delphi studies, either for precision of estimates or for tests of the significance of mean or median differences between two or more forecasts. Consensus and precision are implied from suggestive graphs, not from standard errors of estimates. With small samples and large dispersions, many forecasts do not differ significantly from one another, but are shown to do so by implication if not by explicit statement.

> B4.4 The test documentation should state clearly what interpretations are intended for each subscore as well as for the total test. ESSENTIAL (p. 12).

This standard is especially pertinent to Delphi studies where forecasts are made on a broad and diverse target area. Each forecast should be individually and separately tested for dispersion of consensus,

systematic correlations with other items, and for significance of forecasted differences against other items, as is done with quantitative scores in conventional questionnaire item analyses (Anastasi, 1968).

The author has never seen the full three-dimensional matrix of items versus panelists versus rounds analyzed by a common statistical vehicle, such as analysis of variance, to test for main and interaction effects. Nor are items compared for homogeneity of variance, linearity, and type of empirical frequency distributions for applying such tests. With small samples, interquartile Delphi graphs are no substitute for rigorous statistical testing of individual items and item subsets. This is not a pedantic frill--differential statistical reliability requires differential interpretation of findings.

Except for a study by Derian and Morize (1973), the author has not seen a factor analysis of Delphi items, also part of the standard repertoire in test item analysis. Factor analysis is valuable for pruning out redundant items that are highly intercorrelated, or "saying the same thing" by eliciting the same response from subjects. This type of item "padding" is thus hidden from the end user who interprets results at face value.

If these interpretive standards were respected, quantitative Delphi findings would not be presented in simplistic, descriptive form to potential users. They would then not be taken at face value by users who are unaware of statistical and sampling limitations.

3.3. Empirical Validity

The next items are drawn from the "Validity" section of the APA manual. The keynote standard for this section follows:

> C1. Test documentation should report the validity of the test for each type of inference for which it is recommended. If its validity for some suggested interpretation has not been investigated, that fact should be made clear. ESSENTIAL (p. 15).

This standard provides obvious protection for potential users of test results by requiring the test publisher to indicate whether his test rests on his (vested) opinion (face validity), indirect validity

(e.g., correlations with related areas), or more direct forms of validity testing (e.g., empirical experimentation or real-world performance measurement). With Delphi, panel opinion is reported, with little or no subsequent effort to test results against actual or related events (except for a small number of studies discussed later in this report). The results are usually simply aggregations of iterative opinions. For example, Gordon and Helmer (1964) went no further than to show medians and quartiles and some descriptive scatterplots for their classic forecasting study, and Nanus, Wooten and Borko (1973) simply show frequency distributions and list some percentages for quantitative results in their study of the social impact of multinational computers. Measures of central tendency are put forth, however, as systematic and concurred forecasts of specified events by experts.

The Delphi method typically measures very small sample attitudes toward future events at a given time. It does not measure the events themselves, nor does it incorporate systematic hypotheses and empirical feedback from such events. The leap from raw opinion to future events under these conditions is strictly an act of faith.

The next selected standard is found under "Content Validity." It refers to item definition and item sampling.

> C3. If a test performance is to be interpreted as a sample of performance or a definition of performance in some universe of situations, the test documentation should indicate clearly what universe is represented and how adequate is the sampling. ESSENTIAL (p. 15).

When an area of inquiry has been selected for a Delphi study, as a first step in determining content validity, has the area been adequately formulated and defined? We rarely find systematic reviews of application literature in Delphi studies leading to a careful, state-of-the-art definition of the target domain. Such reviews should extract the best of precursor studies and define basic assumptions and bounds of the inquiry. We often encounter an amorphous socio-technological area (e.g., scientific advances, quality of life, etc.) where the universe of situations may be virtually indistinguishable from future society broadly considered.

The second step in determining content validity is demonstrating that the selected items comprising the questionnaire represent a systematic sampling of key elements of the target area of inquiry. If a particular problem area has been chosen for a Delphi forecast, has a taxonomy been developed for subproblems, embedding situations, resources, and classes of problem-solvers? If so, has it been used as the basis for a representative and comprehensive selection of items?

For example, in using Delphi to forecast computer developments, as was done in Parsons and Williams's widely cited study (1968), content validity preparation would call for a systematic taxonomy of hardware, software, peripheral equipment, communications and applications, perhaps along the lines of the classification scheme used by the Computing Reviews of the Association for Computing Machinery. If the entire computer field is to be covered, or some specified subset, the correspondence between final selected items and the specified area should be spelled out. Such taxonomies, and such accountability in matching items against the target universe, are rarely seen in the Delphi literature.

3.4. Standards for Use of Experts

The next two standards are the only references in the APA manual to the use of experts in test design and analysis. It should come as no surprise that the social sciences have abandoned the use of experts as an integral part of scientific methodology. In test construction and analysis, the role of experts in generating and contributing questionnaire items to the initial item pool is well recognized, and is consistent with current practice. However, the use of experts as the principal and exclusive method for validating tests has been discredited. For example, in World War II, the unreliable "expert" opinions of experienced, professional interviewers were dropped in favor of more effective standardized objective testing procedures (e.g., see Thorndike's account (1949) of the Aviation Psychology Program of the Army Air Forces in World War II).

Another example of the use of experts in the field of economics is revealing. Zarnowitz (1965) studied eight independent forecasts of the

Gross National Product from 1953 to 1963 derived from "expert" opinion. The average observed absolute error for experts was $10 billion, or about 2 percent of the GNP during this period. Zarnowitz found that simple arithmetic extrapolation of the increase occurring in the previous year yielded an average absolute error of $12 billion, effectively the same as the average expert prediction. Zarnowitz conducted studies of other economic indices and obtained similar results.

When we leave the area of short-term forecasting in economics, where extensive baseline statistical indicators are available, and enter the more nebulous areas of psychological and psychiatric diagnosis and prognosis, the record of expert clinical opinion is and has been in a state of disarray. In "The Discontent Explosion in Mental Health," Hersch (1969) explicated the bankruptcy in theory and practice of the unscientific use of clinical experts in empirical research on psychotherapy.

After reviewing some 40 large-scale programs involving man-machine system experimentation in his comprehensive book covering the work in this area since World War II, Parsons (1972) concluded that the reliance of system designers on the opinions and preferences of "so-called expert system operators" is "foolhardy." He pointed out that such experts ". . . may provide suggestive leads, but are not reliable guides, as demonstrated by their repeated disagreement with objective data" (p. 553). These examples are illustrative of the repeated failures and frustrations encountered in the use of experts in diverse social science areas.

> C3.1. When experts have been asked to judge whether items are an appropriate sample of a universe or are correctly scored, the test documentation should describe the relevant professional experience and qualifications of the experts and the directions under which they made their judgments. VERY DESIRABLE (p. 15).

Delphi exercises guarantee anonymity of individual responses to encourage free expression of opinion. Some studies list the names of panelists and, in fewer cases, list their professional affiliations. The author was not able to find any studies listing professional

training and scaled experience levels qualifying each individual as possessing the skills required to meet an objective criterion as an "expert." This "very desirable" standard is effectively neglected in Delphi practice.

> C3.11. When the items are selected by experts, the extent of agreement among independent judgments should be reported. DESIRABLE. (p. 16).

This standard makes an explicit distinction between independent and dependent expert judgment, which gets to the heart of Delphi iteration "with feedback." The first round is basically designed to secure independent expert judgment. The second and successive rounds produce strictly correlated, or biased, judgments. The use of standardized statistical techniques for hypothesis testing based on random sampling assumptions, which may offer no major problems for independent first-round judgments, becomes difficult and problematic in successive rounds--a methodological shortcoming that has apparently not been noticed by Delphi practitioners. All rationalizations about reconsidering, incorporating new information, and converging toward consensus, cannot hide the fact that independent judgment is destroyed once the participant knows how others have responded to each item. If Delphi can make no claims concerning independent expert opinion, does Delphi feedback develop insight into the issues for improved collective judgment in successive rounds?

3.5. Theoretical Standards

The next standard refers to long-term predictions and overlaps substantively with the notion of forecasting.

> C4.41. If a test is recommended for long-term predictions, but comparisons with concurrent criteria only are presented, the test documentation should emphasize that the validity of predictions is undetermined. ESSENTIAL (pp. 17-18).

Delphi practice neglects long-term longitudinal validation, and typically dissociates itself from any systematic comparisons with even

second-string concurrent criteria (e.g., short-term interpretations of long-term trends). Panelists often disagree over what exists "today" and, with rare exceptions, Delphi practitioners make no effort to present panelists with a precise report on "where we are" to establish a baseline for projections into the future. On both counts, for this "essential" standard, Delphi forecasting results should be explicitly presented to potential users as conjectures of undetermined validity.

Delphi practitioners object to this conclusion, pointing out that Delphi has been proven "valid" and "accurate" in a few relatively recent studies involving almanac-type items (Dalkey, 1969) and for relatively short-term predictions (Martino, 1972). Established almanac items (e.g., population of a city, or gross national income at a particular point in time) are not in any substantive way generalizable to long-range forecasts. What they share in common is the trivial property that we all can exercise opinions on each item, hardly a sound basis for generalizing from simple descriptive facts anchored in the past to complex events in the future.

Martino (1972) reports forthcoming work comparing earlier Delphi predictions with outcomes. The original estimates (as in the Gordon-Helmer study, 1964) were derived from pooled respondent opinion, and the outcomes were also determined by pooled opinion. The abuses of such a post hoc subjective approach should be obvious, leaving the central issue of Delphi validity and accuracy unresolved.

The next standard applies to identification of the characteristics of participating panelists.

> C5.2. The validity sample should be described in test documentation in terms of those variables known to be related to the quality tested, such as age, sex, socio-economic status, and level of education. Any selective factor determinng the composition of the sample should be indicated. ESSENTIAL (p. 19).

Delphi studies, having promised anonymity to participants, typically do not report key population characteristics of panelists such as those cited in this standard. Such specification of "expert" samples would permit more effective evaluation of the adequacy of the expert

sample. For example, a long-range forecasting study might benefit from inputs from relatively youthful panelists who are more likely to be living in and directly shaping the world they are forecasting; lower-class or minority members, if the socio-economic items cut across their future; more women panelists, if they are underrepresented (e.g., Dalkey, 1969; Borko, 1970; and Bedford, 1972; have shown systematic quantitative and qualitative differences by sex in Delphi responses); wider geographical distribution of panelists, if they are concentrated in one or two locales; etc. The author has not encountered any studies where panelists have been asked to provide detailed personal data for sampling profiles. Anonymity can still be honored if panelist characteristics are presented as statistical aggregates.

The next standard applies particularly to the pitfalls inherent in the voluntary participation of Delphi panelists.

> C5.3. If the validity sample is made up of records accumulated haphazardly or voluntarily submitted by test users, this fact should be stated in the test documentation, and the test user should be warned that the group is not a systematic or random sample of any specifiable population. Probable selective factors and their presumed influence on test variables should be stated. ESSENTIAL (p. 19).

Panelist dropout is one of the well-known hazards of Delphi experimentation. Delphi dropout rates are probably quite high. Although he cited no empirical data, Martino (1972) asserted that response rates to first-round questionnaires "typically ran 50 percent or less." In the only study the author has been able to find on Delphi dropouts, Bedford (1972) noted that dropouts in a study on home communication services were less motivated to participate in the study (i.e., dropouts responded to fewer questionnaire items), and more significantly, dropouts were considerably more critical of the overall study, the utility of questionnaire items, and of the relative stress placed on various factors such as "lack of concern for sociological and psychological considerations."

There is no question but that some selective factors operate to determine the hard-core group that sticks with the study through all

iterations. The reasons may be positive, such as strong motivation and interest in the target area, or negative, such as a high proportion of personal acquaintances of the director, or of those in his professional circle. Perhaps those who disagree strongly with the design and content of the questionnaire, and those who question initial results (as in Bedford's study), drop out more often than those who have confidence in the study and the procedure, or who play along with minimum effort. To the extent that any systematic panelist sampling effects are known, they should be stated explicitly and taken into account in the evaluation of results. If the original expert sampling is unknown, and if the dropout rate is also unknown, the sample on which the final results are based is doubly suspect. This double indemnity is probably the rule, not the exception, for Delphi studies.

A recent memo sent to me by Brownlee Haydon, illustrates the possibilities of serious social abuse of conventional Delphi in picking a stacked panel of experts in a controversial area with major vested interests.

> "If you are a regular reader of *The New Yorker*, you may already have seen the series entitled "Annals of Industry--Casualties of the Workplace" currently appearing in that magazine. The November 12, 1973 installment describes a classic case of the misuse or perversion of the Delphi process.
>
> "As I read it, Arthur D. Little, Inc. has undertaken for the Department of Health, Education and Welfare (Occupational Safety and Health Administration) to use the Delphi method to arrive at a consensus on the proper level of exposure to asbestos fibres (2, 5, 12, 30 fibres of greater than 5 microns length per cubic centimeter of air) to be established as a government safety standard. What is almost unbelievable is the choice of "experts"--apparently members of the asbestos manufacturing community and their "medical experts" along with a few (too few) independent medical researchers in the field of asbestos-induced cancer!"

Dr. Selikoff was the only member of the "expert health panel" in the Delphi study who ". . . had not been a paid consultant of, or whose investigations into asbestos-related disease had not been supported by, some segment of the asbestos industry . . ." In this *New Yorker* article,

Dr. Selikoff pointed out, "And what's the point of guessing about the biological effects of asbestos when mortality studies of asbestos workers have already shown what the effect has been?" (Brodeur, 1973, p. 172).

The next standard appears under the section concerned with "construct validity," which refers to the interpretation of theoretical constructs on which tests are based. This standard raises the key issue of accountability for the interpretation of Delphi results.

> C7.1. The test documentation should indicate the extent to which the proposed interpretation has been substantiated and should summarize investigation of the hypotheses derived from the theory. ESSENTIAL (p. 23).

This requirement is largely ignored in Delphi practice where a descriptive approach characterizes the presentation of results. The reasons, theories, and hypothetical constructs of expert panelists are covert, rather than overt. Panelists are asked for opinions, and the occasional rationale from panelists is typically very brief, uneven, and often absent in final reports. This haphazard manner of collecting and reporting data underscores the casual opinionative essence of Delphi. There are many levels of opinions ranging from snap judgments to carefully organized and well-defended documentation of positions systematically linked to interpretive concepts of construct validity. Although Delphi practitioners may point out occasional exceptions, snap judgments are apparently the rule for most Delphi questionnaire items, as shown below.

Bedford (1972) appears to be the only investigator who has solicited, classified, and analyzed all panelist comments in his Delphi study on home communications services (for a sample of 1253 responses). His analysis of open-end verbal responses has led him to defect from "traditional Delphi with its heavy emphasis on statistical feedback" toward a structural adversary procedure "stressing the importance of assumptions, qualifications, interpretation of general tiends, and criticism of co-panelist's remarks" (p. 43).

3.6. Questionnaire Reliability

The next section in the APA manual is concerned with test reliability. The first selected standard indicates minimal statistical requirements for reliability reporting.

> D3. Reports of reliability studies should ordinarily be expressed in the test documentation in terms of variances for error components (or their square roots) or standard errors of measurement, or product-moment reliability coefficients. ESSENTIAL (p. 29).

Delphi studies invariably tend to ignore such "essential" considerations of test and item reliability. For example, Sahr (1970) presents some 50 pages filled with descriptive quantitative data comparing three Delphi studies conducted at the Institute for the Future. At no point does he report a single statistic indicating "variances, standard errors of measurement, or product-moment reliability coefficients" required by this standard. Dalkey (1969) has made an initial attempt in this direction by indicating increasing reliability of medians with increasing sample size of panelists--a surprise-free result. (The standard error of measures of central tendency generally vary inversely with the square root of sample size.) He does not present standard errors of medians for individual item results as minimally required by this standard. Dalkey does present split-half (odd-even) reliabilities for some results with coefficients usually varying between .4 and .6. This reported level of reliability is marginal for useful questionnaires. Furthermore, these are for end-results with non-independent or feedback-affected opinions, as discussed earlier. Reliability of first-round results would provide more meaningful coefficients for rigorous statistical testing. Dalkey's attempt to measure reliability is the exception rather than the rule for the descriptive statistics characteristic of the Delphi literature.

For example, Martino (1972) attempts to demonstrate the reliability of Delphi by listing several analogous items in presumably independent studies which resulted in "similar" predictions. No correlation coefficients or other statistical indices are reported, no account is presented of deleted items or discordant items, and no attempt is made

to describe comparability of test conditions for final results. A study by McLoughlin (1969) is cited in which two groups of experts provided independent forecasts for 55 identical questionnaire items. The obtained standard deviation of the differences of the medians between the two groups was 3.54 years for events expected to occur before 1990. Martino concludes that this result shows a "high degree of consistency." On the contrary, assuming a 5 percent level of significance, this finding means that the "true" median forecast falls somewhere between ± 7 years of the obtained forecast (± two standard deviations), which is hardly the basis for claiming a "high degree of consistency." A 95 percent confidence belt of 14 years is not very good for forecasts of events expected to occur with 20 years.

The next standard cited also applies to test reliability, in particular the stability of results.

> D6. Test documentation should indicate to what extent test scores are stable, that is, how nearly constant the scores are likely to be if a test is repeated after time has lapsed. Test documentation should also describe the effect of any such variation on the usefulness of the test. The time interval to be considered depends on the nature of the test and on what interpretation of the test scores is recommended. ESSENTIAL (pp. 30-31).

This "essential" standard says, as applied to Delphi, that the questionnaire should be replicated at a later time on an independent sample of panelists, following original procedures, so that earlier results can be compared with later results to determine test reliability over time. No such replications are reported in the Delphi literature. This type of reliability is especially important for Delphi because the method presumably measures attitudes toward the future which change to a greater or lesser extent with changing conditions and independent panels. The absence of such studies, and the lack of interpretations of the underlying dynamics of attitude changes toward the future, is a major methodological and theoretical shortcoming of Delphi.

Some Delphi proponents object to a study of the underlying dynamics of attitudes toward the future, as distinct from and peripheral to the

domain of Delphi opinion technology. The argument is that opinions are quite different from attitudes, particularly if they are concerned with technical subjects. Such a position reflects the isolation of Delphi from the mainstream of social science. The author concurs with Anastasi (1968) who says "Opinion is sometimes differentiated from attitude, but the proposed distinctions are neither consistent nor logically defensible. More often the two terms are used interchangeably . . ." (p. 480). In this report the two terms are used more or less synonymously.

The validity of any testing instrument cannot be greater than its reliability; that is, a test cannot correlate more highly with any external validation criterion than its correlation with itself (reliability). If Delphi results prove unstable in a given area over the short run, as with attitude fluctuations over time, its value as a prognostic instrument is likely to be worthless over the long run. Longitudinal reliability studies of this type are essential for any defensible use of Delphi or its derivatives.

3.7. Experimental Sampling Standards

The final section of the APA manual covers sampling scales and norms. The next standard overlaps to some extent with prior discussion, but is worth emphasizing.

> F6.11. Norms reported in test documentation should be based on a well-planned sample rather than data collected primarily on the basis of availability. ESSENTIAL (p. 35).

Selection of panelists for Delphi studies tends to reflect expediency rather than a "well-planned sample," particularly when investigators are not accountable for sample specification under the anonymity clause. Heavy Delphi dropout rates can only compound and aggravate this shortcoming.

The next listed standard specifically warns against a standard Delphi practice of developing norms (generalizations) from small samples of panelists.

F6.31. If the sample on which norms are based is small or or otherwise undependable, the user should be cautioned explicitly in the test documentation regarding the possible magnitude of errors arising in interpretation of scores. ESSENTIAL (p. 36).

If Delphi investigators made it common practice to report standard errors of estimates for small samples, it would be apparent to all that higher levels of precision, larger samples, and well-defined samples would be required. This is particularly true where medians are reported rather than means, since the standard error of medians is usually larger than mean errors. It is also the case for forecasts far into the future, where observed dispersions are typically very large, precision poor, and more extensive sampling necessary. Martino (1972), for example, has demonstrated an increasing dispersion of forecasts in many Delphi studies as the expected year of occurrence is farther away.

The next standard describes a practice that has been consistently neglected in the Delphi literature.

F6.4 Test documentation should report whether scores vary for groups differing on age, sex, amount of training, and other equally important variables. ESSENTIAL (p. 36).

The tacit Delphi assumption is that the pooled opinion of experts is better than that of any subgroup of experts. This may or may not be the case for any given area of Delphi inquiry. However, the fact remains that there may be systematic effects related to the kinds of sampling characteristics mentioned in this standard. It behooves the Delphi investigator to test for such effects and to report them, rather than to assume uncritically that the whole is axiomatically better than any of its parts. Dalkey (1969) has demonstrated sex differences for almanac items; Borko (1970) lists substantial sex and professional differences for library and information science research items; and Derian and Morize (1973) show systematic differences between types of medical specialists (e.g., researchers versus clinicians) in medical forecasting.

3.8. Conclusion for Social Science Standards

This concludes the tour through portions of the APA manual of standards relevant to Delphi. It should be abundantly clear that conventional Delphi neglects virtually every major area of professional standards for questionaire design, administration, application, and validation. In no sense is Delphi found to be a serious contender in scientific questionnaire development and in the experimentally controlled and replicable application of questionnaires.

But this is not the whole story by any means. Many key areas remain to round out the picture. Only the methodology common to any questionnaire instrument has been covered. The special characteristics of Delphi remain to be reviewed and evaluated, and this is the task of the next section.

4. DELPHI EVALUATION

The pre-Delphi literature, mentioned earlier, anticipated many of the evaluative problems encountered in the use of opinion to forecast social and technological events. McGregor (1938) and Cantril (1938), from social psychological approaches, found the forecasting process via questionnaire to provide a medium for projecting personal values and attitudes of the respondents. They made no claims for the validity of the technique in forecasting social events, nor for the ability of experts to predict complex social events any better than non-experts. McGregor's conclusion summarizes his findings. "The amount of information possessed by the predictor, and his sophistication or expertness are shown to have little significance in the determination of predictions concerning complex social phenomena. The quality of information as determined by ambiguity and importance is much more decisive" (p. 203). Cantril obtained similar results and concluded that "Whenever the prediction of a social event is based wholly or in part upon an internal frame of reference, objectivity is rare, if not impossible, because of ego-involvement" (p. 388). Both studies illustrate further the difficulties encountered in the use of opinion, expert of otherwise, in predicting events.

Kaplan, Skogstad, and Girshick (1950) summarized the difficulties they encountered, in trying to generalize from their results in social and technological forecasting by questionnaire, as fundamentally a problem of sampling. They concluded that " . . . the most serious question raised by a study of prediction is whether the analysis is made on a statistically stable population. The difficulties are threefold: those concerning the group of predictors, those concerning the questions asked, and those concerning procedure," (p. 108). These authors were skeptical of their findings because of uncontrolled and unknown individual differences between subjects, obvious differences between questionnaire items precluding extrapolations to related areas, and the limitations of the procedure--such as subjective factors in experimenters' judgment, time constraints in selecting items, multiple

choice and probabilistic format of items, and discrepancies between the use of judgment of subjects under experimental conditions compared with the use of experts under more realistic conditions. This was a pivotal study, one that provided key leads for initial Delphi developments. Unfortunately, scientific admonitions concerning statistical representativeness and experimental rigor, as we have seen in the previous section, were disregarded by the Delphi originators.

The critical literature on Delphi is uneven and sparse. Quinn (1971) has described limitations of forecasting in general that apply to Delphi, including such factors as surprise events, inadequate or biased data, and unpredictable interactions. Pill (1971) explores various limitations of Delphi and, in connection with its reliance on human intuition, suggests that ". . . perhaps the Delphi technique should be less allied with science than with metaphysics" (p. 61). Milkovich, Annoni, and Mahoney (1972) emphasize the loss of valuable data because Delphi participants are not allowed to interact directly. Weaver (1969, 1970) suggests that Delphi pays inadequate attention to psychological values and attitudes toward the future. (See Fishbein, 1967, for a comprehensive introduction to the methodological literature on attitude testing.) Morris (1971) has criticized Delphi for not capitalizing on the extensive mathematical literature on the theory of subjective probabilities (e.g., Bayesian analysis). In the previous section, we have shown that this criticism applies not only to advanced probabilistic analyses, but also to elementary statistical treatment of raw Delphi data.

Derian and Morize (1973) criticize conventional Delphi for taking the central tendency of pooled opinion at face value as a best estimate of expert opinion. Through the use of factor analysis of Delphi participants in their study, they found subgroups of experts clustering together with consistent opinions. They recommend analyses of subgroups rather than composite consensus.

> "However, rather than the consensus itself which only expresses the average opinion of the group, the knowledge of the structure of the answers and motivations specific to the different subgroups constituting the panel of experts

> can be extremely useful to the decision maker. Thus, in the case of the artificial heart project, it is not unimportant to be able to assess for instance the differences in points of view between research specialists, clinicians and surgeons."

As mentioned earlier, Bedford (1972) found so many shortcomings in conventional Delphi, that he developed an independent technique called SPRITE: Sequential Polling and Review of Interacting Teams of Experts. In a comparative Delphi study on future home communication services, he found no consistent statistical differences in forecasting results between housewives and experts, and he found the qualitative responses more useful than the quantitative results. This led Bedford to drop the traditional Delphi emphasis on consensus, and move toward "controlled conflict" between contrasting groups, and to drop statistical feedback in favor of qualitative arguments. SPRITE is an example of non-conventional Delphi.

Weaver (1972) has probably contributed the most extensive critical review of Delphi uncovered in our survey of the literature. He asserts that the "vast majority" of Delphi studies "tend to be uncritical" and "promotional." He believes that "Delphi panels cater to the power structure" (p. 21). Delphi studies reviewed "suffer from technical limitations" subjects to experimenter bias in collating and summarizing responses, subjectivity, lack of alternatives, and no checks on wording or order of items. Weaver asserts that "There is serious sterility in the process of summarizing mass information into narrowly terse statements. There is a serious absence of any effort to probe beneath the surface for explanations" (p. 21).

In discussing needed changes in Delphi, Weaver makes several recommendations. He suggests a shift away from mere description of events to explaining events. He would drop anonymity, statistical feedback of dates and probabilities, and "consensus forcing procedures." He questions the notion "that convergence improves the accuracy of a forecast." Weaver would add face-to-face interaction and direct confrontation to ensure exchange of assumptions, arguments, and conclusions, and cites an example of such an exercise conducted at the International

Adult Education Seminar, at Syracuse University. Weaver believes that the elimination of anonymity and statistical feedback, and the introduction of face-to-face confrontation, still represents a recognizable variant of Delphi. It seems to this reviewer that with the rejection of the three "quintessential" elements of conventional Delphi (anonymity, iteration, and statistical feedback) any resemblance between Weaver's recommended interactive group process and conventional Delphi is strictly coincidental. Weaver's recommended approach closely resembles Heller's method of "group feedback analysis" (1969) which was developed independently of Delphi.

In his summary, Weaver asserts that

> "At present Delphi forecasts come up short because there is little emphasis on the grounds or arguments which might convince policy-makers of the forecasts' reasonableness. There are insufficient procedures to distinguish hope from likelihood. Delphi at present can render no rigorous distinction between reasonable judgment and mere guessing; nor does it clearly distinguish priority and value statements from rational arguments, nor feelings of confidence and desirability from statements of probability."

Weaver concludes by urging his recommended changes in conventional Delphi and by stressing its value as an educational and heuristic tool, as distinguised from forecasting.

The author generated a list of advantages and disadvantages of Delphi in his review of the literature, as a preparatory exercise to develop a data base for this critique. The disadvantages soon vastly outstripped the advantages. Approximately 200 negative criticisms were compiled. These were arrayed as ten key questions which are presented below.

The advantages of conventional Delphi, at least in this reviewer's estimation, are primarily concerned with low cost, versatile application to virtually any area where "experts" can be found, ease of administration, minimal time and effort on the part of the director and panelists, and the simplicity, popularity, and directness of the method. However, these and related advantages are characteristically obtained by unwarranted assumptions in method and approach and by seriously

compromising the reliability, validity, and integrity of final results. Such advantages are inconsequential if the conventional Delphi concept, method, and results are inherently untrustworthy.

The ten key questions for conventional Delphi are these:

1. Is the Delphi concept of the expert and its claim to represent valid expert opinion scientifically tenable, or is it overstated?
2. Are Delphi claims of the superiority of group over individual opinion, and of the superiority of remote and private opinion over face-to-face encounter, meaningful and valid generalizations?
3. Is Delphi consensus authentic or specious consensus?
4. Are Delphi questions, particularly forecasting questions, precise and meaningful?
5. Are Delphi responses precise and unambiguous?
6. Are Delphi results meaningful and unambiguous?
7. Is Delphi primarily concerned with collections of snap judgment opinions of polled individuals from unknown samples, or is it concerned with coherent predictions, analyses, or forecasts of operationally defined and systematically studied behaviors or events?
8. Does Delphi anonymity reinforce scientific accountability or unaccountability in method and findings?
9. Does Delphi systematically encourage or discourage the adversary process and exploratory thinking?
10. Does Delphi represent a critical tradition, or is it uncritically isolated from the mainstream of scientific questionnaire development and behavioral experimentation, and does Delphi set a desirable or an undesirable precedent for interdisciplinary science in the professional planning and policy studies community?

Each of these questions is discussed in the ten sections that follow. The answers to these questions, based on this analysis, are found in the eleventh and concluding section.

4.1. Delphi Experts

As emphasized earlier, it is almost impossible to find current psychometric or social science literature on "experts." For example, the author was not able to find any continuing, systematic studies on experts in the recent *Psychological Abstracts* (except for highly specialized applications in legal testimony and clinical diagnosis), in Berelson and Steiner's (1964) inventory of findings on human behavior, in the United Nation's *Encyclopedia of Social Science* (Gould and Kolb, 1964), in the *International Encyclopedia of Social Science* (Sills, 1968), or in many other social science texts that he has examined. Sole reliance on the use of expert opinion for scientific validation has long been discredited. There is a very extensive literature on psychometric scales for judgments, attitudes, and opinions for a variety of tests (e.g., see Anastasi, 1968), and for specified subject populations, but not for "experts."

In assembling a relatively small group of experts, typical of Delphi, the director is tempted to select panelists he knows, or colleagues recommended by his acquaintances. Such selection is tempting because it is easier and faster, with fewer rejections. Perhaps the fastest way to discourage a Delphi study is for the director to fight uphill against a high dropout rate from panelists. The resulting sample of "experts" is likely to include people with similar backgrounds and interests, who think along similar lines. Such groups may also tend to comprise an elite with a vested interest in promoting the area under Delphi investigation. Expert panels are often selected from accessible experts, and this accessibility is largely covert. Delphi reports characteristically offer little or no information about panelist selection, and provide no safeguards against such abuses.

Top names in the field under investigation lend prestige to the Delphi study. The inclusion of prestigious individuals acts as a magnet to attract others less prestigious. However, the prestige personalities may be counterproductive--the younger and more obscure panelists may be more highly motivated to work harder at the questionnaire and provide more carefully considered responses. There is always the choice between the older, established professional versus the young

Turk. Representation of the entire spectrum is probably better than taking sides, at least to help assure more diversified opinion. Turoff (1971) and Martino (1972), alarmed by uncontrolled panelist dropout rates, and concerned with the need for higher levels of panelist motivation and more carefully reasoned responses, recommend budgetary provision for honoraria for panelist time and effort.

The use of experts leads to a serious technical limitation of the Delphi questionnaire--the fallacy of the halo effect, in this case the expert halo effect. This is the tendency of respondents to be unduly influenced by any favorable or unfavorable characteristic of the questionnaire which colors and contaminates their judgment. For example, a highly desired technological event may systematically receive more optimistic forecasts than a neutral event.

Delphi is enmeshed in a pervasive expert halo effect. The director, the panelists, and the users of Delphi results tend to place excessive credence on the opinions of "experts." Panelists bask under the warm glow of a kind of mutual admiration society. The director has the prestige of pooled authority behind his study, and the uncritical user is more likely to feel snug and secure under the protective wing of an impressive phalanx of experts.

The result of the expert halo effect for Delphi is to make no one accountable. The director merely reports expert opinion objectively, according to prescribed procedure; he is not responsible or liable for outcomes. The panelist obligingly follows the ritual, protected at all points by faceless anonymity. The user can always claim that he was simply following the best advice available, and that he is not responsible for what the experts say. Everyone has an out, no one needs to take any serious risks, and no one is ultimately accountable. With so much to gain, so little to invest at such low risk, no wonder the method is so popular. The Delphi belief structure is psychologically held together by the cementing influence of the expert halo effect.

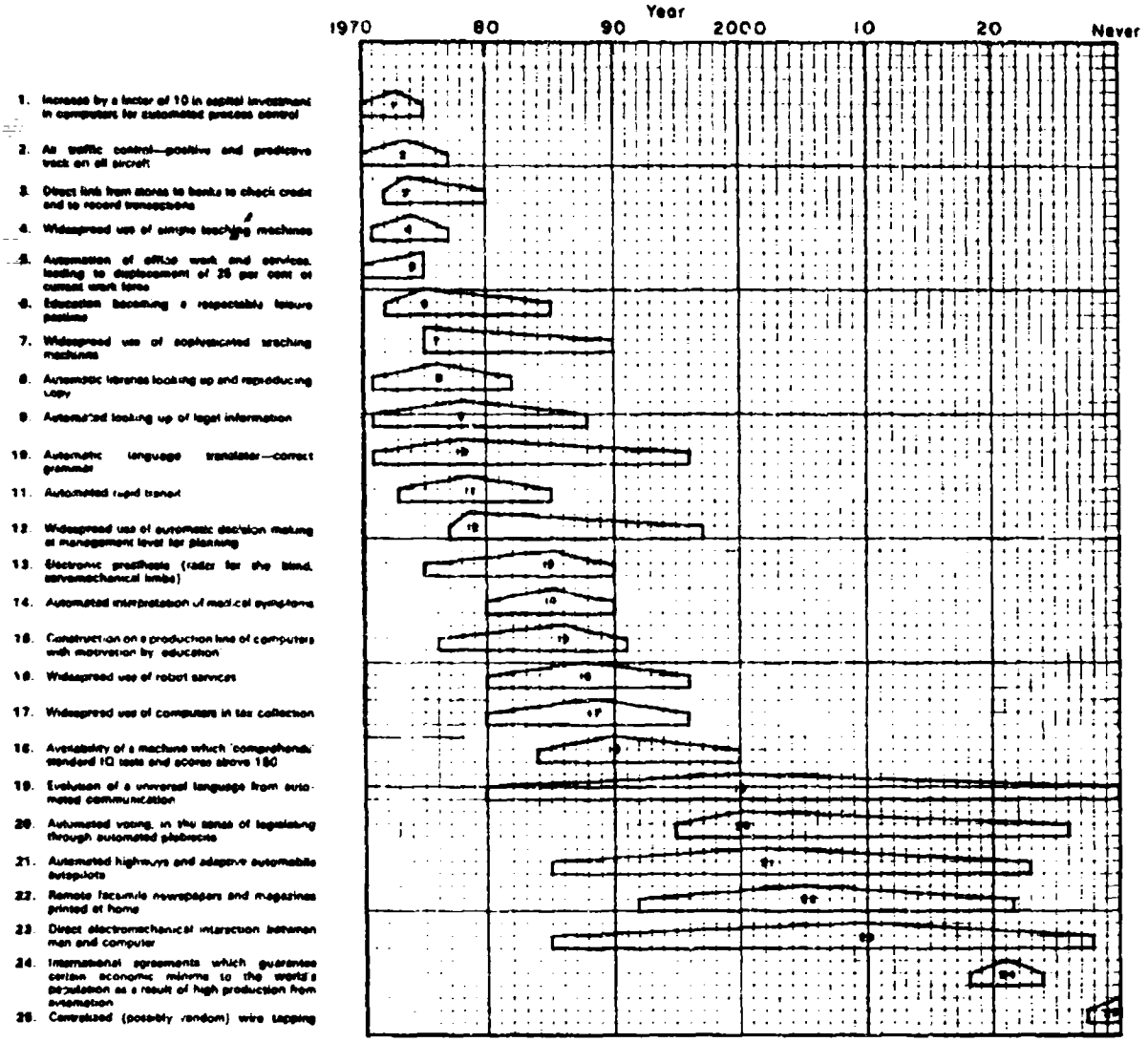
A tacit, largely unchallenged assumption of Delphi is that authentic experts do in fact exist for predicting the extremely complex socio-economic-technological events so common in Delphi questionnaires. Closer scrutiny reveals this to be wishful thinking. Many of these

events are initial forays into unknown areas requiring unknown skills, hence, unknown "experts." Even if such events are understood to some extent, they typically presuppose a fantastic array of real, not shallow, skills in diverse and far-ranging fields such as economics, public policy, esoteric technologies, individual and group psychology, law, medicine, etc., which is simply beyond the ken of any living mortal. When we match predictions of complex sets of social events against "experts," we get something like the fabled blind men examining the Indian elephant. If we think of experts as *idiots savants*, we suddenly avoid the trap of the expert halo effect.

Another central postulate in the Delphi epistemology of experts is that they will in fact provide significantly better and substantially different responses than nonexperts. Practically every Delphi practitioner asserts that Delphi outputs are only as good as the expert inputs, admonishing us with the GIGO principle (garbage in/garbage out).

Suppose, however, that it can be proven that any informed group of individuals in the object area of inquiry can provide individual and group Delphi opinions essentially indistinguishable from those of the experts. It would follow, then, that Delphi results merely represent informed opinion rather than expert opinion.

Personal experience with graduate student predictors brought this potential expert fallacy to the author's attention. In connection with a graduate-level course on computers and society, the author asked his students to give their independent estimates of expected order of occurrence of each of the events in automation (computer technology) and general scientific advances originally investigated by Gordon and Helmer in their 1964 Delphi study. (See Fig. 1 for results with automation items.) After the students ranked the listed events they were told the "true" ranks listed by the experts in the original study, and calculated a Spearman rank coefficient (product-moment correlation of ranks). This provided each student with a correlation coefficient comparing his first-round estimates with the medians of the "experts." Over the years, we have consistently found median rank-correlations for classes of about a dozen students at about .70 for both areas for



TECHNOLOGICAL PROGRESS in automation as predicted by a panel of experts has been obtained by investigators at the RAND Corporation using the Delphi technique. The length of each bar represents various estimates put forward by the 'middle half' of the panel. In each case one quarter—the 'lower quartile'—proposed dates earlier than that at which the bar begins and another quarter—the 'upper quartile'—give dates beyond that marking the end of the bar. Each bar has a peak value which represents the median date estimated

Fig. 1: Delphi Results for Progress in Automation

(Adapted from Gordon and Helmer, 1964.)

first-round estimates. These results are roughly equivalent to the upper levels of reliability for Delphi judgments described earlier from Dalkey (1969).

In a nutshell, "informed" graduate students provided essentially the same forecasts as "experts." The students did have the advantage of making their predictions some six years later than the experts in the original study, items were not presented randomly, there was no iteration with feedback, and standardized instructions were not rigorously observed in these informal classroom exercises. If this general equivalence holds under controlled experimental conditions, anyone with some professional training in broad target fields could play the Delphi game and it wouldn't make any difference in the results.

The tests using graduate students were not conducted as rigorous experiments and the results have not been reported in the literature. No claims are made for the validity of the findings--they are presented here to point up a central hypothesis. Some critical experimental studies comparing experts with less informed individuals and with non-experts have been performed in the Delphi literature and in precursor studies. This is a central empirical question that can be very easily tested.

At the beginning of this section, the studies of Cantril (1938) and McGregor (1938) were cited. In these studies, the expertness of the forecaster was shown to have little or no significance in the determination of predictions of complex social events. More precisely, no statistically significant differences in such predictions were found between students and teachers, laymen and professionals, in tests which involved a combined respondent sample of over 600 subjects. Predictions were demonstrably linked to values and attitudes toward the subject matter.

Kaplan, Skogstad, and Girshick (1950) applied a forecasting questionnaire on 152 social and technological events to 26 subjects representing the entire spectrum from senior professional to layman. Part of the study involved administration of a general knowledge paper-and-pencil test on "Current Social Problems" and "Science." The better-informed subjects (upper half) performed only slightly better than the

less informed subjects (lower half)--average accuracy scores for short-term predictions were 56 percent and 50 percent, respectively. This result is in the expected direction, but is not statistically significant with respect to a test for the mean difference between proportions for this sample. Further, the amount of the difference, as indicated by the authors, is not substantial. Thus, these pre-Delphi studies indicate that expertise maker either no difference, or only a trivial difference, in forecasting a variety of social and technological events.

Much the same results occur with Delphi studies. In Campbell's doctoral dissertation on forecasting short-term economic indicators (1966), level of expertise was tested in terms of self-confidence ratings. He correlated these ratings for each item against forecasting accuracy and found the results did not differ significantly from a median correlation of zero. Campbell concluded that "Selecting the most self-confident members of a group, based on the five-point or the group self-confidence scales, was not an effective means of identifying the most accurate forecasters" (p. 112).

Campbell had additional information for a further test of the relation of expertise to accuracy in forecasting. Of the two seminar groups tested, one group was older and more experienced in professional economic forecasting than the other. The more experienced group did obtain accurate median forecasts more often than the less experienced group in a paired-comparison test, but the results were not statistically significant for Delphi and non-Delphi groups matched against each other for 16 economic indicators. (Because Campbell did not report statistical comparisons, the author applied the non-parametric sign test used by Campbell in similar comparisons and obtained confirmation of the null hypothesis for Delphi and non-Delphi groups.) The pooled results showed 20 more accurate forecasts for the more expert group, 10 for the less expert group, and two ties, which meets a 10 percent level of significance.

Dalkey (1969), also using self-confidence ratings of expertness for each item, was able to compare those "more expert" against those "less expert" for almanac-types questions. "The basic hypothesis being

tested was that a subgroup of more knowledgeable individuals could be selected in terms of their self-rating, and that this group in general would be more accurate than the total group. In every case this hypothesis was not confirmed" (p. 68).

In a subsequent almanac-item study, Dalkey, Brown, and Cochran (1969) did find that ". . . significant improvements in accuracy of group estimates can be obtained with proper use of self-ratings" (p. v.) Close examination of "proper use" reveals rather arbitrary ex post facto statistical verifications that have dubious generality for other studies (e.g., at least seven subjects in high and low subgroups, with no overlap in self-ratings between subgroups, which eliminated many of the subgroups). Such arbitrary ad hoc statistical procedures capitalize on chance fluctuations in the experimental sample. A more appropriate statistic would include all data, such as a correlation coefficient showing both the statistical significance and strength of the association between self-ratings and accuracy.

Bedford's study (1972) is probably the most relevant to the issue at point--are there demonstrable forecasting differences between experts and non-experts? Bedford matched a group of 25 housewives against a group of 26 experts in "communications, consumer behavior, sociology, and futurism generally" in a two-round Delphi study on "The Future of Communications Services in the Home." Bedford found, using a long and extensive questionnaire, that "There were remarkably few differences between the experts and the housewives on the panel" (p. 1). His results support the contention that level of expertise makes little difference in exploratory socio-economic forecasts.

Similar results were obtained by Reisman, Mantel, Dean, and Eisenberg (1969) in a comparative Delphi study. Evaluative ratings of laymen correlated highly with ratings of experts for 250 social service packages handled by the agencies of the Jewish Community Federation of Cleveland. These results also tend to support the hypothesis that opinions for evaluative social areas of inquiry tend to be independent of level of expertise.

What is the box score for the null hypothesis that there are no demonstrable differences between predictions of experts and non-experts

for socio-economic-technological events? The McGregor (1938) and Cantril (1938) studies unequivocally indicate that such differences do not exist for complex social events impacting on personal values. The Bell Canada study by Bedford (1972) indicates that no demonstrable differences were shown between housewives and experts for socio-technological developments. Campbell's analysis of self-confidence ratings also supports the null hypothesis, in that no correlation was obtained with accuracy of short-term economic forecasts. Dalkey's 1969 study showed no differences in almanac items estimates with respect to ratings of self-confidence. Reisman et al. (1969) showed similar responses from laymen and experts in evaluations of social services. These studies collectively indicate that it doesn't make any difference how expert the respondent is, or how confident he feels about his opinion, when forecasting or estimating a wide variety of social, economic, and technological phenomena.

Studies that show some differences in responses between different levels of expertise are marginal at best. The Kaplan, Skogstad, and Girshick study showed a statistically non-significant trend in the correct direction with more "knowledgeable" subjects contributing more accurate short-term forecasts. Campbell's data (1966) also showed a statistically non-significant trend in the expected direction with his more experienced group tending to give more accurate forecasts than the less experienced group. The Dalkey, Brown, and Cochran study (1969) showed statistically marginal results in the expected direction for self-confidence ratings.

If Delphi investigators cannot demonstrate statistically significant and substantial differences between experts and non-experts, then it must be concluded that the Delphi emphasis on the use of experts is misplaced. Available experimental data indicate that this conclusion is probably the most accurate generalization for most Delphi applications. If statistically significant but low-order correlations are found, the expert concept is only marginal, and virtually worthless from a practical point of view. The above experimental data indicate that this might be the case in a small proportion of well-defined and highly specialized applications. If significant and substantial

differences are found, a stronger case may be made for Delphi expert opinion for the target area of inquiry. The above experimental data offer no evidence of substantial differences between experts and non-experts.

Looking back at the central issue of the Delphi concept and use of experts as discussed in this section, we find the following shortcomings:

- The concept of expert is virtually meaningless in experiments dealing with complex social phenomena.
- Sole or primary reliance on expert opinion in the social sciences has long been discredited and now has no serious advocates.
- Anonymous panels chosen in unspecified ways enhance the possibilities for contaminated, elitist "expert" samples.
- There exists an uncontrolled and unknown expert halo effect in Delphi contributing to expert oversell. Collective expert opinion directly reinforces unaccountability for Delphi results for all concerned--the director, panelists, and users.
- Experts and non-experts consistently give indistinguishable responses in forecasting or evaluating social phenomena impacting on common values and attitudes.
- There is no explicit matching of skills required by Delphi questions against objectively measureable skills of the panelists.

The difficulties associated with the Delphi concept of "expert" does not and should not imply that all and any use of experts is necessarily bankrupt. The originators of Delphi should be credited with clearly sensing and trying to respond to strong social demand for exploiting expert opinion more effectively. For example, in a survey of 65 corporations, Hayden (1970) found that 69 percent used diverse

expert panel consensus techniques, and of these, 26 percent used Delphi. This example is probably indicative of the widespread informal and formal use of "experts" throughout society. The proper use of expert talent remains a major problem of our time. We know precious little about the dynamics, the use, and the abuse of experts in our society. Substantive treatment of this problem, however, is beyond the scope of this analysis.

4.2. Face-to-Face Confrontation Versus Private Opinion

Much of the popularity and acceptance of Delphi rests on the claim of the superiority of group over individual opinions, and the preferability of private opinion over face-to-face confrontation. Martino (1972), for example, flatly asserts "It should be remembered that Delphi represents a distinct improvement over either individual experts or face-to-face panels" (p. 27).

Democratic process rests on the secret ballot where voting is performed in private. Group opinion is a time-honored corrective against individual excesses. And how many of us have either been bullied in heated group exchanges or have bullied others when we had the opportunity? Besides, who wants to take the time and effort to travel to a meeting and listen to every panelist defend his expertise to the rest of the group? A quick and incisive statement of the issues on paper and an equally quick indication of individual opinion, also on paper and in the familiar privacy of your own office, as advocated by Delphi, has almost irresistible practical appeal as a sensible and cost-effective solution to the problem of sampling expert opinion.

On the other hand, each of us can probably recount numerous examples where individuals were more effective than groups in arriving at informed opinion; where confrontation clarified the issues and made honest communication possible; where introversion and isolation led to unfortunate aberrations of opinion and outlook.

The experimental data comparing individual and group performance offer no convincing conclusions on either side of these broad issues, although the literature extends over many decades. After reviewing the early literature in this area (1920-1957), Lorge, Fox, Davitz, and

Brenner (1958) indicate that superiority of the group or the individual is relative to stipulated experimental tasks and conditions, varies enormously with individual differences, and is shot through with methodological difficulties in generalizing from experimental to real-world situations. In a more recent review of the experimental literature, Maier (1967) concludes that the comparative effectiveness of individuals versus groups varies widely and depends upon the tradeoff of the assets and liabilities of both approaches in the unique applied setting. He emphasizes the crucial role played by experienced group leaders acting as neutral facilitators in achieving successful group outcomes.

If we look for Delphi studies comparing groups and individuals we find a near vacuum. Dalkey (1969) compared face-to-face with anonymous Delphi interaction for the almanac-type items mentioned earlier. He found a tendency toward more accurate opinion in the anonymous setting, a statistically non-significant tendency. Dalkey's procedure involved picking group "leaders" randomly, which flies in the face of effective group procedure and effectively stacks the odds against successful group interaction. Farquhar (1970) compared group versus anonymous Delphi interaction for a complex software estimation task and consistently obtained substantially better results in the face-to-face group.

Campbell's dissertation (1966) is frequently cited by Delphi proponents as definitive evidence of the superiority of Delphi group opinion compared with face-to-face confrontation in traditional expert panels. Campbell worked out a careful experimental design as far as subject sampling is concerned, randomly assigning graduate student participants to experimental Delphi panels and control confrontation groups (which he called "uncontrolled-interaction groups"). His criterion measure consisted of accuracy in forecasting 16 short-term statistical economic indicators; a flaw in this part of his study is that these 16 measures are only partially independent, which vitiates the integrity of statistical tests based on assumptions of independence. Campbell used nonparametric statistics in comparing median forecasting performance of his experimental (Delphi) and control groups (confrontation), and apparently demonstrated statistically significantly better forecasting in his two matched Delphi groups.

His conclusion, however, is based on a straw-man type of comparison, similar in certain respects to the token conventional group structure used by Dalkey (1969), mentioned earlier. Campbell's control groups were leaderless, and remained leaderless, which undoubtedly led to considerable floundering and non-mission posturing and competition. The simple institution of an elected chairman to organize each group, who would identify with the problem, as occurs in conventional committees, might have altered results substantially. The confrontation groups were force-fitted into a Delphi-type format to make quantitative forecasts more directly comparable. For example, meetings were kept within fixed periods of time, whether or not the group wanted such a procedure, with one meeting corresponding to each round of the Delphi panels; discussion of each economic indicator was also pegged to a fixed period of time, regardless of success or failure in achieving closure or consensus; and each meeting required open individual polling for statistical comparability of estimates between experimental and control groups, whether or not the group wanted to follow such a procedure. These procrustean constraints break most of the rules for professional or enlightened group problem solving. The oppressiveness of these artificial confrontation groups may have undermined group motivation and morale to the point where the meetings became counterproductive, and the comparison spurious. Accordingly, Campbell's study can not be viewed as a serious comparison of the effectiveness of Delphi and conventional panels for his criterion measures.

The results raise additional methodological problems. Campbell did not compare the forecasting results of both types of groups against trend extrapolations of his selected economic time series, even though these series were available on a quarterly basis. It may be that simple arithmetic extrapolation (as mentioned earlier in connection with Zarnowitz's critical review of expert economic forecasting), or perhaps more sophisticated multiple regression analyses, might provide results as good or better than those obtained with expert groups. Finally, quarterly forecasting is hardly a criterion vehicle for an expert panel when reliable and extensive baseline statistical data are available for fine-grain, short-term trend forecasting. Any generalization from such results would have to be limited to very short-range forecasting.

The alleged superiority of anonymous Delphi opinion over face-to-face opinion, and its converse, are unprovable general propositions. They can not be proved or disproved, in general, because the propositions are amorphous stereotypes and are not amenable to scientific testing unless they are operationally defined. Once such definition is applied to limited concrete situations, one approach may prove more effective than another, both approaches may be more powerful than either alone, or the two approaches may be so close as to not make much practical difference. Investigators should be more interested in a flexible eclectic approach that freely capitalizes on the best of both worlds than in identifying with a ritualized approach on either side. In any case, the Delphi claim that pooled group "expert" opinion is more effective than individual opinion, and that anonymous interaction is more effective than direct confrontation, cannot be sustained.

4.3. Delphi Consensus

The goal of the Delphi procedure is to arrive at a meeting of the minds, consensus among the experts. The position taken here is that the Delphi procedure arrives at such a consensus by feeding back the "correct" answer, by rewarding conformity and effectively penalizing individuality, and by proffering non-independent iterative results as authentic expert consensus. Authentic consensus refers to group agreement reached as a result of mutual education through increased information and the adversary process, which leads to improved understanding and insight into the issues; it does not refer to changes of opinion associated primarily or exclusively with bandwagon statistical feedback.

It was stated earlier, in connection with the APA professional standards for soliciting judgment, expert or otherwise, with a standardized instrument, that the judgment should be independent. The first Delphi round represents independent opinion, whereas succeeding rounds are strictly correlated. First-round results of "experts" may contain a range of responses up to four orders of magnitude for some types of quantitative estimates (see Dalkey, 1969, and Baran, 1971), which are hardly publishable as "consensus." Raw-score frequency

distributions are so highly skewed that logarithmic transformations are often required to approximate normal distributions. Perhaps this is why most Delphi investigators do not report first-round dispersions. (Borko, 1970, provides an exception to this rule.)

An example of the logarithmic-range of first-round dispersions for some types of Delphi estimates is provided by Baran (1971) in an illustrative appendix of his report for "Cashless-Society Transactions." This item refers to cost and marketing estimates of hard-copy recording of financial transactions with update balance in computer memory. The first round showed a range of $.01 to $100 for average dollar value of a transaction (10,000:1), 5 percent to 90 percent market penetration five years after mass introduction of this service, and a range of 0 percent to 100 percent for percentage of this service that home subscribers would be expected to pay. The inkblot nature of such future projections speaks for itself.

Now, in succeeding rounds, do the panelists really think through their positions and work toward authentic consistency of opinion, or are they effectively pressured into conformity? Dalkey (1969) has indicated that statistical feedback alone (group medians for each item) is as effective in obtaining consensus as statistical feedback with adversary rationale for responses. Once the panelist knows the median for a problematic item, he has in a very real sense been given the "correct" answer to the item. Panelists are quite aware that median responses (or some other measure of central tendency) are offered as best estimates for questionnaire items in the final results.

Social psychologists have long been aware of powerful tendencies for individuals to conform to group opinion in relatively unstructured situations, particularly if the motivation level is not high (Stogdill, 1959, and Berelson and Steiner, 1964). The "autokinetic" effect is a striking example of this tendency (Sherif, 1936). Place an individual or a group of people in a completely darkened room with a single, fixed point of light. The light will appear to drift randomly with a displacement as high as 20 degrees, because of the absence of a visual frame of reference. (Astronomers were the first to notice and study the autokinetic effect.) Ask the subjects in such a room to estimate

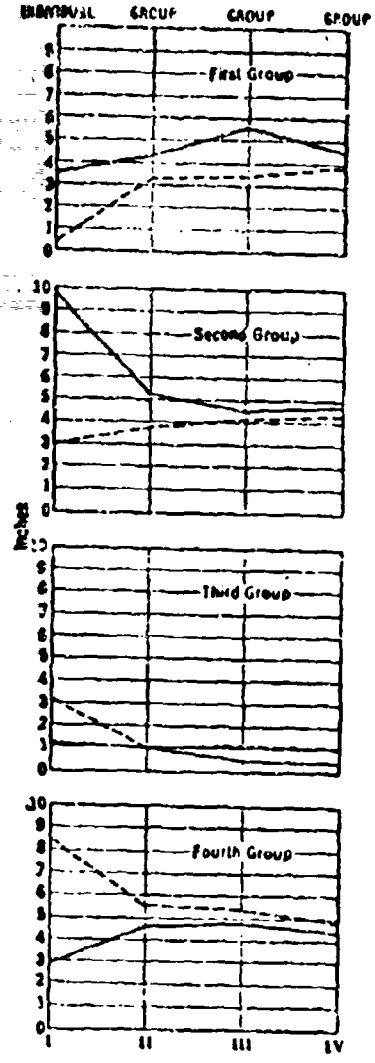
the direction and amount of perceived movement of the light. Initial random judgments soon converge closely around the group norm after a few rounds of group opinion. Group suggestion provides the "correct" answer to an inherently ambiguous situation. Consensus is specious.

Figure 2 shows some of Sherif's experimental results with the autokinetic effect. The first session involved individuals alone reporting observed deviations of the pinpoint of light in a completely darkened room. The ordinates in Fig. 2 represent median deviations in inches, the abcissas represent successive sessions (equivalent to Delphi rounds). The second, third, and fourth rounds were group sessions where each individual had an opportunity to hear the deviations reported by others. Note that the individual median deviations rapidly converge to a group norm by the fourth round for groups of two or three subjects.

The analogy with Delphi is startling. Convergence of medians is greatest with initial feedback of group opinion, and is effectively achieved in three to four rounds. Delphi investigators typically reach the point of diminishing returns at about three or four rounds as far as measurable convergence of opinion is concerned. When we couple Sherif's results with Dalkey's assertion that statistical response alone is the most effective way to achieve consensus (without verbal feedback) we have the artifact of autokinetic consensus (group suggestion) explaining Delphi consensus. Sherif ran many variations of the autokinetic effect demonstrating easily *manipulated* shifts in subjects' opinions in *any* desired direction by suggestion from the experimenter (e.g., "you are underestimating light movement") or from other authority figures, such as group leaders. The uncontrolled, arbitrary introduction of selected verbal feedback by the Delphi director can with corresponding ease shift opinions in desired directions.

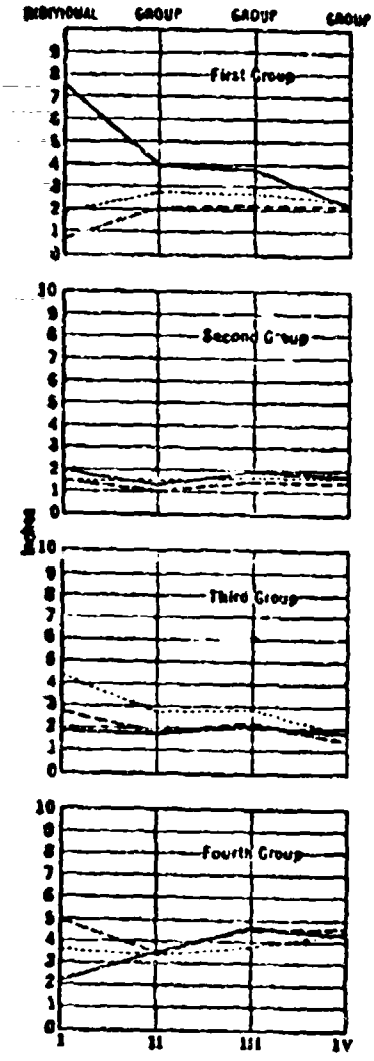
The Delphi technique thus deliberately manipulates responses toward minimum dispersion of opinion in the name of consensus. The presentation of median opinions (after the first round) and the coercion toward conformity are reassuringly represented to all as reasoned

Individual
Median Deviations in
Groups of Two Subjects



Successive Rounds

Individual
Median Deviations in
Groups of Three Subjects



Successive Rounds

*Figure 2 Specious Consensus: Autokinetic Convergence

*Adapted from Sherif, 1936.

——— Subject 1
– – – – Subject 2
-------- Subject 3

consensus. By the time the third or fourth round occurs, the holdout individualist responses pose the threat of yet another tedious run through the same items, and even die-hards are inclined to yield to save everyone the dreary routine of another round. Martino (1972) recommends, ". . . in many cases, there is no advantage in going beyond two rounds" (p. 27).

In passing, it should be noted that the term "panelist" is a misnomer in the Delphi context. Panelists usually communicate directly and exchange opinion with each other, primarily in a face-to-face setting. With Delphi, we have respondents, not panelists, because communication is strictly with the questionnaire, not with other people. Moreover, all responses are filtered through the intermediary of the Delphi director or his representative before reaching anyone else. There is no interactive discourse deserving of the name "panel" in Delphi procedures. Respondents really represent a non-communicating non-group, linked primarily by remote statistical feedback.

Delphi consensus is suspect from still another viewpoint. The first-round items are quite different when they are accompanied by statistical and verbal feedback provided by the director in succeeding rounds. Once the information accompanying an item is altered, it is literally a different item. Just as minor rewording can change a questionnaire item enormously, so does Delphi "feedback" change the item in uncontrolled and unknown ways. How can medians and dispersions be compared, and consensus claimed, if items are noncomparable from round to round?

The social implications of specious consensus are enormous. Variations of similar iterative query techniques, with conformist-reinforced feedback, provide almost unlimited possibilities for shaping and manipulating public opinion via the interactive communications media of the future.

4.4. Delphi Questionnaire Items

The basic criticism leveled against Delphi questionnaire items is that they are, by and large, unavoidably amorphous. More specifically,

complex future events (and value judgments) do not lend themselves to clear and unambiguous description in typical one-sentence Delphi questionnaire format (e.g., see the automation items from the Gordon-Helmer study in Fig. 1). Instead we find vague, generalized descriptions of future events, permitting the respondent to project any one of a large number of possible scenarios as his particular interpretation of that event. Delphi asks panelists about event-stereotypes, and panelists respond with stereotype estimates. Delphi verbal responses, when they occur, are typically vague and sweeping descriptions, slogans, or simplistic statements.

More thoughtful and careful Delphi investigators attempt to qualify forecasts by identifying percentages of specific respondent populations and by associating probability estimates with predicitons. Such attempts, although in the right direction, are no substitute for precisely defined, detailed scenarios for each item where a host of assumptions specifying the "event" are made explicit. The questionnaire format does not lend itself to such presentation.

For example, the Delphi inquiry might be concerned, as in Baran's study (1971), with the "Potential Market Demand for Two-Way Information Service to the Home." Baran had to leave vast areas unspecified in asking panelists when such services were likely to be available and how much they would cost the consumer. These unspecified areas included the configuration of hardware, software, and communications; the nature of federal, state, and local regulation of such mass computer services; the mix of public and private support of the information services considered; very brief general descriptions of the 30 information services (typically one paragraph); no indication of how the public will be taught to use such services; and many other socio-economic-technological areas impacting directly on these services. Baran's study is probably one of the best available in the Delphi literature, featuring extensive use of computer support, and a rational quantitative and probabilistic cost format for couching questionnaire items. But even with all these precautions, which are considerably more than are encountered in the typical Delphi study, the items incorporate vast areas of ambiguity and represent an array of possible

specific events "fitting" into each item. (Recall the "cashless-society transaction" item, cited previously.) As presently practiced, Delphi is--in many respects--a psychological projective technique for future inkblots.

In his recent experimentation with Delphi procedures in the field of drug abuse, Thompson (1973) underscores the top-priority need for extensive pre-testing, and the great difficulties encountered in developing reliable and useful Delphi questionnaires.

> "The most challenging aspect of future applications of Delphi techniques to the drug field will almost certainly be the design of a cohesive set of questionnaire items that are both well-posed and useful to the decisionmaker. On the one hand, it became apparent during the study that developing concise questions which will be given similar interpretations by all respondents will inevitably involve substantial pre-testing. The usual difficulties in questionnaire design are compounded in the context of drug abuse by disagreement over underlying assumptions, and by the absence of an agreed-upon vocabulary."

The psychological literature on attitude and opinion testing has described an instructive historical process that appears to have gone unnoticed by the Delphi community. After an initial era of free-wheeling, broad-gauged questionnaires on attitudes, covering almost anything of interest, the evolutionary trend has been toward highly specialized attitude and opinion instruments concerned with investigation of specific issues in depth (e.g., Anastasi, 1968). In the Delphi context, this means that single items are often of sufficient complexity to warrant construction of a complete questionnaire dedicated exclusively to that item, exploring major implications and aspects, to better reveal the constellation of opinions to which it gives rise. This permits the development and test of theory to explain and enhance understanding of the item or issue in question.

4.5. Delphi Responses

If Delphi questions are ambiguous, then Delphi responses are also ambiguous. The structure and dynamics of Delphi responses contribute

to compounding the ambiguity. This is due to several factors. Among the most important is the pitfall of inviting snap answers to amorphous questions.

Delphi investigators rarely analyze and report the effort panelists put into responding to their questionnaires. Solicitation requests, and instructions accompanying Delphi questionnaires, typically assure the panelist that the forms can be quickly and easily filled out. If not, the investigator runs the risk of massive dropout rates, as occurred with a 19 percent first-round response in Kochman's study (1968). Assurances are often provided that forms should not normally require more than about an hour of the panelist's time for each round. Martino (1972), for example, recommends an upper limit of 25 items for Delphi questionnaires.

In the absence of information on panelist effort, the author timed his own responses for two Delphi studies in which he was a panelist. The results showed great variation from item to item, with an average of one minute per item where few comments were written, to an average of about two minutes per item for heavily annotated justifications of responses. The typical sequence would be to read the item; think quickly about key critical factors influencing the forecast; peg the crucial factor, if any, or fall back on a general stereotype, if available; get a crude estimate of its occurrence; and fit the gross estimate into the questionnaire scale. The average one minute pass per item was armchair, top-of-the-head opinion, for strictly ball-park estimates. The average two-minute-per-item session involved spending almost three hours on a lengthy form, with many annotations, which was as much as the author was willing to contribute. Although this experience is not necessarily representative, it is difficult to conceive average speeds very much faster than a fraction of one minute per item or, at the other extreme, Delphi questionnaires taking more than half a day of the respondent's time for a single round--even if the data are collected in the costly form of a personal interview.

The author is participating in a Delphi study being conducted by Bell Canada to assess future home communication service trends. Response times for the first round have been carefully recorded. The

Delphi director indicated in the instructions that the 207-item questionnaire should not require more than an hour to answer. The author took 65 minutes. The average time per question was 19 seconds. The range of response times, for different groups of items (usually one page per group), varied from 40 seconds per question to 9 seconds. The pacing mechanism was "fastest possible reading speed for comprehension and instant response." The author indicated to the Delphi director that he had no confidence in such free-association judgments.

An "analysis" averaging one minute or less for complex forecasts is merely a snap judgment, experts notwithstanding. The results are free-association attitudes toward the future, not analyses of future events. We also tend to get order-of-magnitude responses, particularly for quantitative data. As mentioned earlier, this is particularly apparent in first-round results.

Responses tend to represent stereotyped thinking, as illustrated by the following comments taken from the Nanus, Wooten, Borko study (1973): "this technology is essentially here already, so I'll forecast early" or "utopian dreamwork, so I'll forecast never," or "costs are much to high--appear later," or "no one cares, the public won't buy it," or "this is a trivial advance," or "this will kill scientific progress," or "people will rebel against this invasion of privacy." This is not to deprecate the talent and experience of experts, but most human beings, when placed in a situation where they are regarded as experts, accountable to no one, and expected to provide quick answers to complicated questions, are quite likely to lean very hard on stereotypes.

The hypothesis has been advanced in various contexts in previous sections that Delphi forecasting is a form of psychological projection of inkblots of the future. Anyone familiar with psychological projective techniques, such as the Rorschach inkblot test and the Thematic Apperception Test, will appreciate the fundamental basis of such techniques--there are as many "correct" answers as there are respondents. The respondent projects his own emotions, needs, attitudes, imagination, experience, stereotypes, and personal problems into the amorphous stimulus situation, modulated by distinguishable cultural factors

related to age and education (Sackman, 1952). We saw that group conformity factors prevailed in the autokinetic situation studied by Sherif (1936). It has been pointed out earlier that the typical single-sentence questionnaire format for Delphi is such an unstructured stimulus that it amounts to an inkblot scenario of the future. We have noted that one minute per response is typical of Delphi exercises; it is also characteristic of psychological inkblot tests where subjects are urged to free-associate to amorphous stimuli. Rorschach investigators have collected thousands of responses to the standardized set of ten inkblots (originated by Hermann Rorschach), and have tallied responses and published statistical norms of popular and unusual responses. They do not assert that the most popular responses (e.g., butterfly, dancing girls) are "true" or "accurate" responses. By the same token, Delphi investigators have no basis for equating popularity with validity for their "inkblot" results.

Delphi proponents object to this characterization and insist that the statement of Delphi questions in objective, quantitative format yields objective, quantitative results, not amorphous personality projections on arbitrary inkblots. We have already cited order of magnitude, log-normal dispersions possible for first-round quantitative estimates. At this point, additional experimental evidence as to the underlying dynamics of such dispersion is presented in support of the hypothesis that Delphi forecasts are often no more than "inkblot" projections of the future.

McGregor conducted a large-scale study (1938) of psychological determinants of individual predictions of social events. One part of his study was concerned with the impact of the type of information given to respondents when they were asked to make their forecasts. Table 2, reproduced from his study, shows results obtained under three conditions in response to the request to estimate the size of the Communist Party in the United States for the next year (1936 at the time of the study). The three conditions include 1) no information, 2) correct information (e.g., 35,100 members in 1935 with official figures for prior years, and 3) incorrect information where the true figures were multiplied by five. There were two groups of subjects, 246 in the

first and 376 in the second. The first group effectively went through two rounds of this question, initially with no information (first column of results in Table 2), and later with incorrect information (third column in Table 2).

The data in Table 2 reveal several notable results. First, the estimates of the uninformed group were too high by an order of magnitude (first column showing an interquartile range in the hundreds of thousands). Popularity had no relation to accuracy. Note the tendency toward order-of-magnitude clusters at tens of thousands, hundreds of thousands, and millions. With accurate baseline statistics, as expected, the forecasts in the second column of Table 2 were less variable and far more accurate, more like simple short-term trend extrapolation. The third column in Table 2, roughly analogous to a second-round Delphi with feedback, shows how easy it is to manipulate quantitative individual and group opinion to cluster closely around erroneous or misleading data if the situation is sufficiently unstructured. The point of this example is that the inkblot hypothesis applies to quantitative as well as qualitative data for unstructured situations such as quantitative Delphi forecasts of complex social phenomena.

We have already discussed the contamination of opinion with "feedback" in second and successive rounds, and we need not dwell any longer on the well-established finding that individuals tend to shift their expectations to conform to overt group norms, such as a Delphi median issuing from experts. The iterated expert response to each Delphi item is thus built on snap judgment on the first round, followed by various forms of overt and covert conformist pressure in succeeding rounds.

4.6. Delphi Results

Delphi group results are merely collections of results for individual questionnaire items. The items are rarely linked together with theoretical or systematic constructs; this potpourri contributes to a mixed bag of findings. As mentioned earlier, item reliability and item validity are typically ignored, making it easier for the uninformed user to accept results at face value. Standard errors of

Table 2

ESTIMATES OF THE MEMBERSHIP OF THE COMMUNIST PARTY OF THE UNITED STATES UNDER THREE CONDITIONS[a]

Inter-Quartile Range	Without Knowledge N = 246 100,000 to 1,000,000	With Correct Knowledge N = 376 33,000 to 38,000	With Incorrect Knowledge N = 246 160,000 to 180,000
Percent predicting between 30,000 and 40,000	5	83	0.8
Percent predicting 50,000 or less	21	97	2
Percent predicting between 150,000 and 200,000	7	0.3	76
Percent predicting 1,000,000 or more	30	0	0
Mean prediction	?[b]	35,100	172,000

SOURCE: Reproduced from McGregor, 1938.

[a](1) Without knowledge of the membership for past years, (2) with correct knowledge of the membership for the past five years, and (3) with incorrect knowledge (figures five times too large) of the membership for the past five years.

[b]The calculation of a mean from these estimates would have been a meaningless operation because the distribution revealed no central tendency. There were "clusters" of estimates: (1) below 25,000, (2) around 100,000, (3) around 500,000, and (4) between 1 and 5 million.

estimates for forecasts and ratings are usually absent in the Delphi tradition, thus giving the final results an aura of precision. Interquartile graphs knock out 50 percent of the sample and the embarassingly long tails of the extreme non-conformists (e.g., see Fig. 1). The final report may include a few anecdotal comments on selected items, but rarely any connected discourse on controversial interpretations. There may also be a few caveats on the limitations of the study.

The presentation of raw frequency distributions of aggregate opinion generates serious problems for the user in the interpretation of the results. Many forecasters may not differ significantly from one another with respect to the null hypothesis for mean or median differences. Many items may be highly redundant, with similar or indistinguishable results, reflecting a pervasive halo effect. The antidote is to test for differences between items in a systematic analysis of variance for items, subjects, and rounds, as mentioned earlier, to determine main and interaction effects. Redundant items can be discovered through this technique, or through factor analysis of items, as is routine in conventional analysis of questionnaire items. In an unusual exception to standard Delphi neglect of statistical analysis, Dalkey and Rourke (1971) used a type of cluster analysis for quality of life indicators which reduced a very large number of initial raw items to a much smaller number of relatively independent composites or factors. We have no idea how rampant item redundancy and associated halo effects are in the results of the Delphi literature at large, especially with the characteristic absence of techniques equivalent to item factor analyses. It is easier, cheaper, and perhaps more impressive to present the naive user with unprocessed raw data resting on face validity.

After perfunctory qualifications, the investigator makes it quite clear that the experts have pronounced concurred judgment. This is the trump card in the Delphi game. With the apparent tacit agreement not to criticize other Delphi investigations, the results tend to remain unchallenged.

4.7. Delphi Epistemology

A fundamental epistemological confusion exists between Delphi method and Delphi results. Practitioners claim that the end result of a Delphi study is a series of expert forecasts of future events, or more broadly, concurred estimates of whatever social attribute is under study. Prior discussion has provided grounds for a very different interpretation of Delphi results.

Delphi items are typically broad, amorphous classes of events, not precisely defined empirical occurrences. Delphi forecasts are opinions about such broad classes of events, not systematic, documented predictions of such events. These opinions are typically snap judgments frequently based on free-association stereotypes. Consensus for such opinion tends to be manipulated consensus to minimize dispersion of opinion. Further, the universe from which items are sampled is typically disregarded and unknown, as are the identity and qualifications of the expert panelists.

Orthodox Delphi epistemology holds that the result of this type of polling procedure produces reasonable and useful forecasts of object events. This worthy goal is not attained. The Delphi process produces manipulated convergence of opinion reflecting ephemeral attitudes of very small samples of unknown individuals. More precisely, Delphi produces transient attitudes about the future, which is quite different from systematic predictions of the future. The epistemological confusion arises from focusing on Delphi results and naively taking them at face value as expert predictions of the future, rather than looking at the underpinning method which reveals Delphi as an attitude polling technique dealing in snap judgments of ill-defined issues.

There is a closely related epistemological issue concerned with Delphi validity--the so-called accuracy of Delphi predictions. Observers continue to say "How accurate is Delphi?" "Prove that its accuracy is better or worse than other techniques." These questions presuppose a scientifically replicable calendar/stopwatch concept of forecasting validity where an impartial observer with a stopwatch waits for the objective event to happen, clocks it, and records the time and date of occurrence. This is fine for simple, unambiguous,

factual items such as "When will man first land and walk on the moon?" The calendar/stopwatch concept can not be applied to such items as "widespread use of robot services, automated rapid transit, use of computers in tax collection, automated legal information retrieval, etc." (see Fig. 1, once more). There are as many scenarios for each of these items as there are respondents. How can anyone validate the truth or falsity of an inkblot of the future?

The way out for some Delphi investigators is to ask the experts at a later date whether the forecasts have materialized (Martino, 1972). However, this results in another opinion poll, or opinion validated by opinion, not an objective assessment of external events. This amounts to bootstrap validation--Delphi validating itself. Such studies, if conducted rigorously, would provide an indication of longitudinal test-retest reliability (correlation of Delphi with itself over time), not an indication of application validity which requires correlation against an external criterion.

In limited application areas, such as immediate or very short-range forecasts (excluding the questionable applications to almanac items), Delphi accuracy can be measured. Farquhar's study (1970) of the estimation of software manpower requirements, previously mentioned, is one example. Delphi performed very poorly when compared with face-to-face groups in this case. Delphi forecasting of well-defined short-term economic indices based on Campbell's (1966) doctoral dissertation at UCLA, was not shown to differ substantively from simple extrapolation of short-term time series data. In 1952, Helmer published the results of a Delphi study predicting the results of the 1952 presidential contest between Eisenhower and Stevenson. After four rounds, the seven panelists converged on Stevenson as the winner.

Even this very limited and inconclusive sample of studies indicates that Delphi results will often be untrustworthy, and will vary enormously between, and even within, object problems or application areas, reflecting differences in experimenters, "experts" selected, particularly with the ground rules and baseline data made available to them, and numerous other methodological issues. If these frequently untrustworthy and highly variable results over various application

areas are characteristic of relatively immediate or short-range estimates, it can be fairly confidently inferred that medium and longer-range results will, by and large, be even more variable in reliability and validity. Recall, for example, Martino's findings (1972, cited previously) that the standard deviations of forecasts for independent studies are consistently highly and positively correlated with median expected values of occurrence. (For example, scan the widths of the interquartile "houses" in Fig. 1 with increasing estimated median of occurrence.) Put simply, the farther in the future an event is expected to occur, the more uncertain the prediction is likely to be.

A concrete example illustrates the scope and magnitude of the inkblot problem for Delphi accuracy or validity. Suppose the Delphi questionnaire asks "When will mass information utilities become commonplace?" The range of "correct" answers for this item, depending upon the scenario projected by the respondent, can literally vary from the Western Renaissance to beyond the year 2000. If "mass information utility" is interpreted to mean mass-produced books, then the answer is somewhere in the sixteenth century, after the introduction and spread of Gutenberg's printing press. If interactive long-distance conversation is the preferred scenario of the respondent, then the advent of the telephone in the late nineteenth century is the answer. If the expert interprets the item to mean mass electronic broadcasting, he would identify the radio as the source, and opt for the early 1920's. Another expert might interpret the item as meaning audiovisual broadcasting, and list the 1950's for the mass use of television. Another respondent might interpret the item as involving mass computerized transmission of information, and indicate the mid-1970's as the point where computerized information may greatly exceed non-computerized information over various transmission media. If the item were interpreted as two-way, interactive computer services in the home, as in Baran's (1971) study, the respondent might pick the 1980's. A cosmopolitan expert, accepting the same scenario, but thinking of popular use throughout the entire industrialized world, would place his prediction in the next century. Although this illustration is deliberately extreme, the central point should be quite clear--the Delphi

questionnaire format does not lend itself to scientifically objective and externally verifiable statements of future events.

In contrast to the above example, consider more conventional technological forecasting studies under the sponsorship of NASA (Feldman, 1965) and the Air Force in Project Forecast (Amsler and Newton, 1963). In both of these studies, the authors assembled extensive data on engineering characteristics for specialized forecasting targets--communication satellite output devices (Feldman), and multipurpose long endurance aircraft (Amsler and Newton). Qualifying specifications and assumptions were spelled out, technical baselines were carefully defined and established, and most likely technological developments were projected. Most results were expressed quantitatively, often in graphic format. The key difference between these results and conventional Delphi results lies in the rigorous technical framework in which the forecasts were embedded. These NASA and Air Force examples illustrate initial steps in the direction of operationally defined predictions essential for scientifically verifiable forecasts.

Thus, when someone asks "How accurate are Delphi results?" the answer should be that "Accuracy can not be measured for most Delphi items, because changing attitudes and opinions on amorphous issues are not true or false and do not have specific dates at which they occur." Asking for proof or disproof of Delphi accuracy amounts to giving Delphi credit for generating results capable of proof--a property that conventional Delphi, as currently practiced, does not possess.

There is nothing inherently wrong with studying and learning more about opinions concerning the future. Such knowledge is crucial to any intelligent appraisal of the future. But we should not confuse such opinion with seriously considered, qualified and documented predictions of well-defined future developments. Attitudes and opinions change and fresh sampling in real time is needed to track such changes. And the sampling must be explicit in terms of subject populations if any systematic inferences are to be made.

4.8. Anonymity and Accountability

The anonymity of Delphi experts serves the dual purpose of attracting expert panelists by guaranteeing protection against individual accountability, and projecting an inviting image of a kind of permissive brainstorming where "anything goes" to help "cream off" the best the experts have to offer. The panelists are assured full protection against any invasion of privacy. When coupled with the blandishments of joining the inner circle of eminent experts, the combination is hard to resist. But few have realized that the price of such inducements is abandonment of accountability, and may promote elitist vested interests.

Under a "no disclosure of names" policy, no individual is accountable for either his own responses or for group Delphi results. As pointed out earlier, Delphi embodies circular buck-passing. The director reports group opinion following an objective ritual; the results are not his personal opinion. Each panelist is faceless in any of the results, and can always blame nameless others for any findings he dislikes. The consumer of Delphi gets his low-cost preview of the future, and can claim he had nothing to do with the final results.

Directors should be accountable for all flaws in the method, and for implicitly or explicitly overstating the value and significance of potentially misleading final results. Panelists should be accountable for unwittingly lending the authority of their reputations and their support to demonstrably unreliable and invalid short-cuts to the future. Individual and institutional users should be accountable for funding and popularizing such studies, and for accepting Delphi forecasts at face value.

Elitist tendencies are strongly reinforced not because of any diabolical plot on the part of Delphi investigators, but for the more mundane and more compelling reason that it is a lot easier and faster to assemble colleagues, acquaintances, or second-order recommended acquaintances for the expert panel.

A major attraction of Delphi for busy researchers of all callings is that it is cheap and easy, as well as a relatively painless and well-protected technique. A study can be conducted and a paper produced

with relatively small effort. Martino (1972) claims that ". . . a planning factor of two professional manhours per panelist per questionnaire is a fair approximation to the workload which will be required" (p. 60).

The questionnaire is quite likely to represent many aspects of the work done by these experts, almost by definition. Chances are that while such panelists will have much to disagree over, most will be interested in promoting the image, value, and particularly the future of their field. Big developments will then be perceived as occurring early and making large impacts on society. For example, Nanus, Wooten and Borko (1973), in their Delphi study on the social implications of multinational computer systems, admit that their sample of 56 "eminent" panelists, typically active in various aspects of this field, were probably biased to some extent toward promoting the importance and enhancing the image of multinational computer systems. There is no malevolent design or covert collusion in such opinions, merely self-aggrandizement and self-interest.

Controlled experiments soliciting opinions from contrasting or even antagonistic groups (expert or otherwise) are likely to produce quite different results. As mentioned earlier, Delphi opinion polls measure attitudes toward future events, not predictions of such events in their own right. As currently practiced, Delphi can easily slant results in the direction of aggrandizing vested interests. With anonymous sampling of "experts," the burden of proof should be on the Delphi investigator to demonstrate that his panel does not represent a narrow elitist circle.

Kopkind (1967), in his widely cited article on "The Future Planners," expressed his concern over futurist elitism. "The danger is that Government and corporate elites will monopolize the business of question-asking, and so manipulate the attitudes of society they are pretending to serve as disinterested technicians" (p. 23).

4.9. Adversary Process

Most Delphi practitioners claim that Delphi is able to go where other investigators fear to tread. Opinion can peer into every nook

and cranny, particularly those inaccessible to conventional techniques, Delphi thus has the advantage of being able to get "there" first, or among the first, and of making early pronouncements concerning new horizons far in the future. This capitalization on novelty is part of the dramatic appeal of Delphi. Plumbing the depths and climbing the heights of the future hold spills, thrills, chills, and some jolts of future shock for everyone.

It would seem plausible that at least until we learn a good deal about any new domain, it should be the object of free inquiry and of very active adversary proceedings. Delphi systematically inhibits the adversary process. This indictment is not in any sense original with the author. As cited in various contexts throughout this critique, variations of this indictment have been made by, Bedford (1972), Milkovich, Annoni and Mahoney (1972), Turoff (1972), and Weaver (1970).

Delphi deliberately factors out face-to-face confrontation, and the adversary process associated with it, as one of its prime philosophical tenets justifying efficient consensus. Arguments are filtered, buffered, and effectively neutralized in Delphi. A panelist can participate without providing any justification for any of his opinions throughout the entire procedure. More conscientious panelists provide occasional brief commentaries.

The real payoff for the Delphi investigator is obtaining maximum consensus from the experts. Interquartile Delphi forecasting graphs, spreading from now to never, are the nemesis of Delphi practitioners. The smaller the spread the more powerful the impact. Real adversary excitement over authentic controversial issues is plainly the enemy of consensus. Boredom and snap responses make for smaller differences and maximum consensus. In many cases, only the outliers have to justify their positions in Delphi iteration; directors make minimal demands on those occupying the middle ground.

By inhibiting the adversary process, Delphi also inhibits open exploration of new domains. Free exploration leads to adversary inquiry and generates new controversy. This can lead to polarization of opinion that undermines consensus in final Delphi results. But it is precisely the new domains that need free exploration and the adversary

process the most. Delphi should be prodding conformers and rewarding outliers to maximize exploration, highlight controversy, and map out the unknown. When we are really ignorant we need all the contrasting viewpoints we can get to encourage free and informed choice.

4.10. Delphi Isolationism

The history of Delphi reveals a highly exploratory and tentative technique that was never validated. Delphi was obviously full of problematic issues and potentially serious flaws, and was treated with some measure of caution and skepticism by its Rand originators before the Gordon-Helmer study (1964) catapulted the technique into international prominence. After that point, the shaky hypotheses on which Delphi rested were apparently transformed into axioms, and Delphi was promoted as an established, proven technique.

Only relatively recently have Dalkey and some of his co-workers made attempts to demonstrate the validity of Delphi, as reviewed in this report, primarily with almanac-type items and non-expert panelists such as college students. These efforts, and spotty returns from a small number of other studies mentioned in this review, provide no scientific validation of Delphi. This history of early experimentation and tardy efforts to assess validity reflects a pattern of isolationism from the mainstream of behavioral research.

Delphi has led a protected existence for the decade it has been actively pursued. From exploratory and tentative beginnings at Rand, it has spread from government to industry and academia, and diversified from scientific and technological forecasting to policy studies and planning, to quality of life assessment, and is being touted as the emerging nexus for human communication and decision making (e.g., Turoff, 1972). Droves of eminent people and experts from all callings have lent their name, time, and effort to hundreds of Delphi investigations. All this, and undoubtedly more to come. Why?

In part, because there has been virtually no critical literature. The roots of this criticism-free development of Delphi are found in two sources--the isolation of Delphi from the mainstream of relevant behavioral science, and the rapid concurrent emergence and growth of

futurism. For various reasons, Delphi originators and subsequent Delphi practitioners have shied away from psychometric and opinion survey specialists who could have professionalized Delphi as an opinion polling technique along the lines previously suggested in connection with the discussion of the APA test manual and social science standards. The existence of this isolation is attested by the fact that there are virtually no listings of Delphi studies in the *Psychological Abstracts* as revealed by our literature review. The proof of this isolation is the disregard and unconcern for professional questionnaire standards in Delphi practice that has been heavily documented in this study.

The reasons for such isolation are not hard to find. The professional standards would immediately transform Delphi from a cheap and easy, short-cut technique to a far more difficult, expensive and time-consuming procedure. Unprepared and untrained Delphi investigators would have to develop new skills in psychometrics, opinion sampling and polling, and experimental design with human subjects, and would lose considerable control over the technique if experts in these skill areas were taken seriously.

Delphi practitioners and many futurists, broadly considered, identify themselves as interdisciplinarians. They sought to enlist the necessary diversity of skills to assessments of the future. This is most commendable if taken seriously. The place to begin, however, is with the disciplines vital to the method. This was never done with Delphi.

Neither the originators of Delphi, nor subsequent practitioners, have been willing to attempt to establish rigorous standards, and to police the Delphi literature by discriminating between better and poorer work. This has contributed to the spate of crude Delphi studies generated by neophytes.

This lack of standards is characteristic of new disciplines going through early growth. Futurism has not been heavily pursued for much more than a decade. Delphi played no small part in getting futurists on the map by dignifying forecasting with its seemingly impressive ritual for obtaining expert consensus. Other methods, such as brainstorming, scenarios, gaming, input-output analyses, contextual mapping,

simulation, and morphological analyses also experienced rapid growth during this period, contributing largely undisciplined exploratory techniques to futures forecasting and planning. Each technique needs adversary checks and balances for healthy growth, and futurism as a whole needs to develop minimal professional standards and a vigorous critical literature representing more authentic interdisciplinary work.

4.11. Results of the Analysis

This portion of the study, concerned with analysis of the specific and unique assumptions and principles of Delphi, as distinct from opinion questionnaires and human experimentation broadly considered, was organized under 10 key questions formulated at the outset. The analysis suggests the following answers to the 10 questions for conventional Delphi:

1. The Delphi concept of the expert, and its claim to represent valid expert opinion, is scientifically untenable and overstated. As summarized by Professor Haythorn, an external technical reviewer of this report, ". . . the procedure by which the selection of subjects occurs is not properly explicated, the exact nature of the panel of experts is often left unspecified, and the implicit assumption that results obtained using conventional Delphi with a panel of experts is better than or different from results that would be obtained using another population has not been empirically established."
2. Delphi claims of the superiority of group over individual opinion, and of the superiority of remote and private opinion over face-to-face encounter, as well as their counterstatements, are unproven generalizations.
3. Delphi consensus is specious consensus. As succinctly stated by Professor Haythorn, ". . . the group process used in Delphi rounds is quite similar to the techniques used in social psychological research to study group conformity, rejection of deviant opinion, and deindividualization, all of which have been shown to be counterproductive with regard to the quality of group decisions."

4. Delphi questions are likely to be vague.
5. Delphi responses are likely to be ambiguous.
6. Delphi results probably represent compounded ambiguity.
7. Delphi is primarily concerned with transient collections of snap judgment opinions of polled individuals from unknown samples, which should not be confused or equated with coherent predictions, analyses, or forecasts of operationally defined and systematically studied behaviors or events.
8. Delphi anonymity reinforces unaccountability in method and findings.
9. Delphi systematically discourages adversary process and inhibits exploratory thinking.
10. Delphi has been characterized by isolation from the mainstream of scientific questionnaire development and behavioral experimentation, and has set an undesirable precedent for interdisciplinary science in the professional planning and policy studies community.

5. EPILOGUE

The following 16 conclusions sum up the evaluation of conventional Delphi in regard to its method and application. This report finds conventional Delphi:

1. Often characterized by crude questionnaire design.
2. Lacking in minimal professional standards for opinion item analyses and pilot testing.
3. Highly vulnerable on its concept of "expert" with unaccountable sampling, and in the selection of panelists, expert or otherwise.
4. Abdicating responsibility for item population sampling in relation to theoretical constructs for the object area of inquiry.
5. Virtually oblivious to reliability measurement and scientific validation of findings.
6. Capitalizing on the fallacy of the expert halo effect.
7. Typically generating snap answers to ambiguous questions representing inkblots of the future.
8. Seriously confusing aggregations of raw opinion with systematic prediction.
9. Capitalizing on forced consensus based on group suggestion.
10. Unwittingly inhibiting individuality and any adversary process by overtly and covertly encouraging conformity and penalizing the dissident.
11. Reinforcing and institutionalizing premature closure, using a highly questionable ritual for conducting opinion studies that tends to inhibit more scientific approaches.
12. Giving an exaggerated illusion of precision, misleading uninformed users of results.
13. Indifferent to and unaware of related techniques and findings in behavioral science in such areas as projective techniques, psychometrics, group problem solving, and experimental design.

14. Producing virtually no serious critical literature to test basic assumptions and alternative hypotheses.
15. Denigrating group and face-to-face discussion, and claiming superiority of anonymous group opinion over competing approaches without supporting proof.
16. Encouraging a short-cut social science method that is lacking in minimum standards of professional accountability.

5.1. Final Evaluative Recommendations

Two alternative final recommendations were considered as conclusions of this evaluation. One was to seek to upgrade Delphi by recommending higher standards, more consistent with scientific method in the collection, analysis, and use of questionnaire data. The other was to conclude that the assumptions and principles on which conventional Delphi is based are so unscientific and inherently misleading that they preclude any attempts to improve the technique. This second alternative was tantamount to a recommendation to drop Delphi completely.

The evidence adduced in this study clearly indicates that the massive liabilities of Delphi, in principle and in practice, outweigh its highly doubtful assets.

As the preferred alternative to conventional Delphi, professionals, funding agencies, and users are urged to work with psychometrically trained social scientists who can apply rigorous questionnaire techniques and scientific human experimentation procedures tailored to their particular needs. It is recommended that conventional Delphi be dropped from institutional, corporate, and government use until its principles, methods, and fundamental applications can be experimentally established as scientifically tenable.

5.2. Beyond Delphi

Some will grant the very shaky opinionative structure of Delphi, and insist that Delphi was never really put forth as science, but merely as a heuristic vehicle for exploring vague and unknown future issues otherwise inaccessible. They might insist that Delphi as an exercise has generated many insights and has been well received.

Jolson and Rossow (1971) have commented on the heuristic value of Delphi in facilitating communication in the corporate environment. Reisman et al. (1969) have noted the communication potentials of Delphi for community participation in evaluating alternative social services. Even as a heuristic exercise, it would be highly advisable to mix iterative polling with varying forms of quantitative and qualitative feedback, personal confrontation where feasible, cultivated development of adversary positions as opposed to consensus, and controlled variations in the type and level of anonymity. As we have seen, there is nothing sacred in the Delphi process--all basic assumptions, particularly in informal exercises, should be systematically challenged, examined, and tested with other eclectic approaches, and tailored to the unique mission and needs of the object problem.

Brainstorming, if done properly, is fun, generates many insights, and can be well received. Advocates of brainstorming no longer present their results as finished products. Practitioners of Delphi publish results in journals, as master's and doctoral dissertations (e.g., Kochman, 1968, Campbell, 1966, and Weaver, 1969) as major corporate reports (e.g., North and Pyke, 1968), as significant social indicators for national and international planning (e.g., Dalkey, Rourke, Lewis and Snyder, 1972, Bjerrum, 1968), as results worthy of weighty consideration--the embodiment of balanced expert opinion.

There is a vast difference between Delphi as an informal forecasting exercise among questionnaire respondents, and Delphi as the authentic embodiment of thoughtfully concurred expert opinion wherever it is applied. Nanus, Wooten, and Borko (1973) in a relatively ambitious Delphi study on the social impact of the multinational computer, make it clear that no claims are made for the reliability or validity of their Delphi results--Delphi was used for strictly exploratory purposes in an uncharted domain, ". . . the authors chose to use Delphi with realization that the results would be more in the nature of a structured 'brainstorming' session with noted thinkers than a scientific exercise in prediction" (p. 11).

The rejection of conventional Delphi recommended here should not in any way be construed as denying the growing and urgent need of society

to learn and understand more about the future. Perhaps the greatest of all human rights is the right to help shape and determine one's own and society's future. We need to know far more about human attitudes toward future developments.

It is to be hoped that forthcoming opinion polls will systematically sample attitudes toward the future from all segments of the population for more effective and more humanistically informed social planning. Delphi, with its exclusive reliance on small coteries of "experts," has unwittingly fostered another form of elitism to set the pace and formulate the pattern for attitudes toward the future.

The originators of Delphi had the right instincts in responding to growing and pressing needs to enlist the active participation of geographically distributed professionals to work in concert assessing unknown and complex problems. Perhaps their most significant insight was the concept of physically distributed teams building a cumulative base of knowledge through the mechanism of temporally spaced interaction and feedback. Although this concept responds to a strongly felt social need, the implementation has been counterproductive. The originators arrived at premature closure along the lines of an iterative ritual producing ambiguous results.

Instead of testing a great variety of flexible alternatives, the method zeroed in on iterative statistical group response. The alternatives could have branched out into structured adversary procedures including dialectical planning (e.g., Mason, 1969), adversary polling between groups with vested interests as in SPRITE (e.g., Bedford, 1972), iterative online teleconferences (e.g., Sackman and Citrenbaum, 1972), and eclectic mixtures of confrontation and isolated responses (e.g., Heller, 1969, Weaver, 1972). All of these areas need vigorous experimental work.

It is beyond the scope of this analysis to enter into a systematic review of areas of inquiry related to Delphi and possible offshoots that might lead to useful advances in method and findings. Suffice it to say that many research opportunities exist for teleconferencing, iterative polling, the analysis of human attitudes toward the future, cooperative problem solving among geographically dispersed individuals,

and the social dynamics of real, not specious consensus which should be based on a profound understanding of the adversary process in its own right.

Consumers of information on the future need far better advice and protection from contributing professionals than they have gotten to date. The future is far too important for the human species to be left to fortune tellers using new versions of old crystal balls. It is time for the oracles to move out and for science to move in.

Preceding page blank

Appendix

SEMI-ANNOTATED DELPHI BIBLIOGRAPHY

This bibliography includes standard and annotated citations to the Delphi literature. Much of this material was assembled by Barbara Quint of the Rand Library staff. Annotations are included as available from our sources; entries are arranged alphabetically by author.

We were greatly aided by Delphi listings made available to us from the following sources:

1. *Delphi and Long-Range Forecasting*, SB-1019, The Rand Corporation, Santa Monica, California, 1972. (Annotations from this source are indicated by (R) at the end of the listing.)
2. *Selected Bibliography on Delphi Literature*, Institute for the Future, Menlo Park, California, 1972.
3. Pill, Juri, "The Delphi method: Substance, context, a critique and an annotated bibliography," *Socioeconomic Planning Science*, Vol. 5, pp. 57-71, 1971. (Annotations from this source are indicated by (P) at the end of the listing.)
4. Turoff, Murray, "Delphi and its potential impact on information systems," *AFIPS Conference Proceedings*, Vol. 39, AFIPS Press, Montvale, New Jersey, pp. 317-326, 1971.
5. *Annotated Delphi Bibliography*, provided by Michael T. Bedford, Bell Canada Business Planning Group, Montreal, Canada, 1973. (Annotations from this source are indicated by (B) at the end of the listing.)
6. A search through various standard indexes in the Rand library.

The Bell Canada bibliography and the Rand bibliography provided the most extensive annotated listings. The primary focus of the Bell Canada entries is on corporate applications of Delphi. These entries include listings and cross-references for corporations using Delphi which are retained for the convenience of the reader. The Rand entries primarily cover the historical and methodological literature.

The accumulated sources should enable the reader to obtain a reasonably balanced picture of the Delphi technique with numerous applications over many areas in a single alphabetical listing.

BIBLIOGRAPHY

ADELSON, M., M. ALKIN, C. CAREY, and O. HELMER
"Planning Education for the Future: Comments on a Pilot Study"
American Behavioral Scientist, Vol. 10, No. 7, 1967.

The character of American education is determined by many related decisions. Improving it will require a broad base of participation within and outside of school systems. This requirement implies a need for generating and disseminating information about education, and for devising procedures for bringing informed judgment to bear on the decision process in a regularized way. It may be as important to improve the decision process in education as to modify any of the specific features of contemporary schooling. The trend toward systematizing or rationalizing the decision process seems promising, although there is a need to avoid centralized control of the process of developing new citizens who are to live in a democratic society. The future role of the federal government in American education is one of the deep residual issues.

AIL

See paper by PACKARD that describes the use of Delphi at AIL for forecasting the development in the LSI chip industry.

ALDERSON, R. C., and W. C. SPROULL
"Requirement Analysis, Need Forecasting, and Technology Planning Using the Honeywell PATTERN Technique"
Technological Forecasting and Social Change, Vol 3, No. 3, 1972, pp. 255-265.

The authors describe the development and use of the Honeywell PATTERN technique--Planning Assistance Through Technical Evaluation of Relevance Numbers. This technique may be regarded as a distant cousin to Delphi since groups of experts are used to develop consensus on the relevance numbers for the projects under consideration. Although the technique was developed for military purposes, the article uses examples of a personal transportation decision and a bio-medical study conducted by Honeywell. (B)

ALLPORT, GORDON
Becoming
Yale University Press, New Haven, 1955.

AMARA, R. C., A. J. LIPINSKI
"Some Views on the Use of Expert Judgment"
Technological Forecasting and Social Change, Vol. 3, No. 3, 1972.

AMENT, R. H.
"Comparison of Delphi Forecasting Studies in 1964 and 1969"
Futures, Vol. 2, No. 1, March 1970. (Also Institute for the Future, P-9.)

AMERICAN PSYCHOLOGICAL ASSOCIATION
Standards for Educational and Psychological Tests and Manuals
Washington, D. C., 1966.

AMSLER, R. C., and J. S. NEWTON
Multipurpose Long Endurance Aircraft (MPLE) Airplane Design Analysis
Northrop Corporation, NOR-63-109, June 1963.

ANASTASI, A.
Psychological Testing
Third Edition, Macmillan, New York, 1968.

A.T.&T.
The Future of the Telephone Industry
Sponsor of the Institute for the Future Study R-20, (see BARAN and LIPINSKY).

A.T.&T., and WESTERN ELECTRIC
Communications Needs of the Seventies and Eighties
December 1971. (Internal document.)
Reports the results of a Delphi study sponsored by the above companies. High priority communications needs are described in 13 categories. These needs are ranked by categories. (B)

AYRES, ROBERT UL
Technological Forecasting and Long-Range Planning
McGraw-Hill Book Company, New York, 1969.

BARAN, P.
Potential Market Demand for Two-Way Information Service to the Home
Institute for the Future, R-26, 1971.

BARAN, P.
The Future of Newsprint
Institute for the Future, September 1971.
McMillan Bloedel Ltd. sponsored this study which was designed to develop a better understanding of the factors that will have a significant impact on the demand for newsprint and newspapers in the next thirty years. The study examines the future of newsprint and substitute media: non-wood fiber paper, magazines and books, electronic systems (TV and CATV). After an examination of the newspaper business, the study examines some broader issues: ecological considerations, international commerce and the U.S. economy. Many charts of the specific results are included. (B)

BARAN, P., and A. J. LIPINSKI
The Future of the Telephone Industry
Institute for the Future, R-20, September 1971.
This A.T.&T. sponsored study examines five main areas that will have significant impact on the future of the communications business. These are: regulation, social change, existing services and networks, new services and networks, labor force and urban change. The five

panels consisted of 210 respondents from inside and outside the Bell System. (B)

BEDFORD, MICHAEL T.
A Technology Assessment of Future Home Communications Services, A Study Proposal
Bell Canada, Business Planning Paper No. 12, May, 1973.

BEDFORD, Michael T.
"The Value of 'Comments' Analysis and an Analysis of SPRITE as a Planning Tool"
Delphi: The Bell Canada Experience, Bell Canada, October 1972.

This article examines the use of the Delphi technique in the corporate environment. The paper indicates that a broad range of Delphi studies have been conducted or sponsored by various corporations in North America and Europe. These studies were conducted by consultants or planning groups in the business firms. Four Delphi planning groups conducted by Bell Canada's Business Planning Group are examined. Results from these studies are illustrated. Several important issues that must be considered when conducting these studies in the corporate environment are considered. These include: (a) should corporations pay for this type of basic research? (b) how can the results from Delphi studies be best utilized in business? (c) misusing Delphi results in business, (d) in-house versus consultant conducted studies, and (3) the proprietary nature of business Delphi study results. The paper concludes with some projections on the future of Delphi in the corporate environment.

BEDFORD, MICHAEL T.
"The Value of Competing Panels of Experts and the Impact of 'Drop-outs' on Delphi Results"
Delphi: The Bell Canada Experience, Bell Canada, October 1972.

The primary objective of the analysis was to determine whether the division of a Delphi panel into fairly distinct types of panelists would increase the efficiency of the technique in terms of information generated. A second objective was to determine whether a number of panelists dropping out of the pa[illegible]re the final questionnaire has a significant effect on the c[illegible]me. It was found that there was very little difference between the statistical responses of the two groups studied (housewives and communications experts) but that the comments and opinions generated were extremely valuable in developing an internally consistent view of the future. The attempt to learn more about the nature of the "drop-out" panelist was unsuccessful due to the small sample of drop-outs in this particular example, but a data base in this area has been formed and will be updated with future study results.

BEDFORD, MICHAEL T.
The Future of Communications Services into the Home
Bell Canada Business Planning, September 1972. (Proprietary)

This study employs a two-panel approach to estimate future acceptance of communications services in the home. A panel of housewives and a panel of "experts" prepared forecasts and were asked to resolve

differences between the groups' views in a following round. Areas studied included Electronic Shopping from the Home, Remote Banking, Electronic Home Security, Electronic Programmed Education in the Home. Ten types of Information Retrieval Services into the Home were also explored. (B)

BELL CANADA, BUSINESS PLANNING GROUP
(Rm. 1105, 620 Belmont, Montreal 101, Quebec)

Sponsor of six Delphi studies into the future of four subject areas: Education, Medicine, Business, and Home Communications. See references by authors Bedford, Day, Doyle, and Goodwill, Feldman, and Goodwill. The Trans Canada Telephone System booklet: "Communications Computers and Canada" also summarizes the results from the first three afore mentioned studies. The Martino article in the *Futurist* (1972), discusses the Business Planning approach to Delphi research and planning in the corporate environment. (B)

BELL CANADA
Delphi: The Bell Canada Experience
October, 1972.

BENDER, A. D., A. E. STRACK, G. W. EBRIGHT, C. von HAUNALTER
"Delphic Study Examines Developments in Medicine"
Futures, June 1969, p. 289.

The authors describe experiences at Smith, Kline and French Laboratories with the conduct of internal and external Delphi studies on future medical developments. Five areas of medicine development are discussed: biomedical research, diagnosis, medical therapy, health care, and medical education. Results in these areas are displayed graphically. A medical scenario of the 1980's is presented. Comparison of medical results from other studies is shown. Opinions on the reasons for differences between the internal and external SKF panels are also offered. (B)

BERELSON, B., and G. A. STEINER
Human Behavior: An Inventory of Scientific Findings
Harcourt, Brace and World, New York, 1964.

BERNSTEIN, G. B.
A Fifteen-Year Forecast of Information Processing Technology
Research and Development Division, Naval Supply Systems Command, January 20, 1969.

BERNSTEIN, G. B., and M. J. CETRON
"SEER: A Delphi Approach Applied to Information Processing"
Technological Forecasting, Vol. 1, No. 1, June 1969.

BJERRUM, C. A.
Forecast 1968-2000 of Computer Developments and Applications
Copenhagen: Parsons and Williams, 1968. Summarized in C. A. Bjerrum, "Forecast of Computer Developments and Applications: 1968-2000," *Futures*, Vol. 1, No. 4, June 1969, pp. 331-338.

BORKO, H.
A Study of the Needs for Research in Library and Information Science Education
Institute of Library Research, University of California, Los Angeles, California, 1970.

BRIGHT, JAMES R.
A Brief Introduction to Technology Forecasting: Concepts and Exercises
Pemaquid Press, Austin, Texas, 1972.

This work book is designed to be used in conjunction with the courses taught at Bright's Industrial Management Center. Chapter 5, "Delphi Studies as an Aid to Corporate Planning" is a detailed analysis of the Delphi study conducted by Ling-Temco-Vaught (LTV). Corporate background data, sample questionnaires, and 84 study forecasts are illustrated. The LTV Delphi experience is also used to provide a data base for the Chapter (II) on Cross Impact Analysis. (B)

BRODEUR, PAUL
"Annals of Industry--Casualties of the Workplace"
New Yorker, November 12, 1973.

BROWN, B.
Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts
The Rand Corporation, P-3925, September 1968.

A description of the Delphi method and some of the areas to which it has and could be applied. Choosing the panel of experts, whose expertise could be decided on various grounds, is one problem of the method; another is the questioning technique itself. Questionnaires have been used in the past; however, this method could become cumbersome. In a few years it should be possible to equip each expert with a console for feeding responses to a computer, which would then compute the group response and feed back the results. Six experiments using the Delphi method have indicated that it may prove useful in military, educational, and business planning, as a tool for forecasting future strategic, economic, or other states. Other possible applications include medical diagnostics and investment counseling. (R)

BROWN, B., S. W. COCHRAN, N. C. DALKEY
The Delphi Method, II: Structure of Experiments
The Rand Corporation, RM-5957-PR, June 1969.

A compilation of the experimental designs, questionnaires and resulting group response data representing the raw materials of a Rand evaluation of Delphi procedures. (Analysis of the data and major conclusions are presented in RM-5888.) The Delphi technique uses an anonymous, orderly program of sequential individual interrogations, with controlled feedback from respondents between interrogations, to elicit and refine group judgments where exact knowledge is unavailable. Ten experiments involved university students as subjects and posed questions of almanac-type information having numerical answers. The overall aim was to explore how groups use incomplete information to arrive at factual conclusions. Different experiments tested different hypotheses.

One was designed to compare the relative accuracy of group answers obtained by the Delphi questionnaire-feedback method with those obtained by a structured, face-to-face discussion. (R)

BROWN, B., and O. HELMER
Improving the Reliability of Estimates Obtained from a Consensus of Experts
The Rand Corporation, P-2986, September 1964. (Also published as appendix to O. Helmer, *Social Technology*, New York, Basic Books, 1966.)

A report on an experiment in the use of expert opinions. The experiment, involving the Delphi technique and the computation of a consensus based on self-appraised competence ratings, is described and its results analyzed. (R)

BUROS, OSCAR KRISEN (Ed.)
The Sixth Mental Measurements Yearbook
Gryphon Press, New Jersey, 1965.

CAMPBELL, R. M.
"The Delphi Technique: Implementation in the Corporate Environment"
Management Services, Vol. 5, No. 6, November-December 1968, pp. 37-42.

This article investigates the possibilities of using Delphi in the formal corporate environment, and points out some of the techniques in this context. The method is explained in the more or less standard manner, and the importance of panel selection and the need for some structure to evaluate and use the results are stressed. Its use is proposed for the development of new products. The article is brief and general. (P)

CAMPBELL, R. M.
A Methodological Study of the Utilization of Experts in Business Forecasting
Ph.D. Thesis, University of California, Los Angeles, 1966.

This thesis examines current (1966) business forecasting techniques, the forms of group forecasting, and develops a number of hypotheses related to the use of Delphi. An experimental set of Delphi studies is developed and conducted to test the hypotheses. The forecast data are 16 economic and business statistical series. After analysis, the author draws a number of conclusions on the Delphi technique. He also suggests a number of marketing applications for the Delphi process. (B)

CANADIAN COMPUTER/COMMUNICATIONS TASK FORCE
Branching Out, Vol. 2
Information Canada, Ottawa, May 1972.

The task force report examines various policy options for the Canadian government in the computer communications field. The second volume contains various appendices. The appendix on Education applications integrates the results of the Bell Canada Education Delphi study with other material (pp. 101-129). (B)

CANTRIL, HADLEY
"The Prediction of Social Events"
Journal of Abnormal and Social Psychology, Vol. 33, 1938, pp. 364-389.

CARSON, ROBERT
Interaction Concepts of Personality
Aldine Publishing Co., Chicago, 1969.

CARTWRIGHT, D., and A. ZANDER
Group Dynamics
Row Peterson, Evanston, Illinois, 1960.

CETRON, M., and D. OVERLY
"Toward a Consensus on the Future"
Innovation, No. 31, New York, May 1972.
This study is an interesting one, even though it is more of a survey than a Delphi study. The authors have gone through their data bank of forecasts, many from Delphi studies, and prepared a questionnaire on 74 future events in two broad categories: "general" events and "technical" events. The first category, business, political, and social changes, has a time frame to 1980. The second category goes to 1980 and 1985. The respondent will be anyone who tears out the questionnaire from the journal and returns it to the authors. Since the journal is business oriented, we can expect that a large number of businessmen will respond. Answers to be published in the future. (B)

CETRON, M., and C. RALPH
Industrial Applications of Technological Forecasting
Wiley-Interscience, Toronto, 1971.
This book is designed to stress the practical applications of T/F in industry. The use of Delphi by several corporations is outlined (often unnamed). These firms include: "A Petrochemicals Research Institute," "Man-Made Fibers Co.," (material shown), Monsanto Corp., Smith, Kline and French (examples), LTV (examples), "Large Copper Company." (B)

COCHRAN, WILLIAM G.
Sampling Techniques
Wiley, New York, 1963.

"Computers in the Crystal Ball"
Science Journal, August 1969, p. 15.
Short reference to the Parsons and Williams Delphi. Two charts from the study showing projections on computer applications and computer development are shown. (B)

CON-FORM
See study described in the LACHMANN article.

CURRILL, D. L.
"Technological Forecasting in Six Major U.K. Companies"
Long Range Planning, March 1972.

Describes a survey by the author of 100 U.K. companies (37 respondents) on technological forecasting. Techniques used by six companies are described. Delphi is used by a "Glass" Company, a "Consumer Goods" Company, 2 "Chemical" Companies, and an "Electrical Engineering" Company. Study results not shown, but the importance of various techniques in the companies is described. (B)

DALKEY, N. C.
"Analyses from a Group Opinion Study"
Futures, Vol. 1, No. 6, December 1969, pp. 541-551.

DALKEY, N. C.
The Delphi Method: An Experimental Study of Group Opinion
The Rand Corporation, RM-5888-PR, June 1969. (See also RM-5957, P-2983, and P-3721.)

A report of results of experimentation on the effectiveness of Delphi procedure, which incorporate anonymous response, iteration and controlled feedback, and statistical group response to elicit and refine group judgments where exact knowledge is unavailable. In spring 1968, Rand conducted ten experiments using over 150 university students. Questions related to almanac-type information. Results showed that controlled feedback, compared with face-to-face discussion, improved the accuracy of group estimates, thus validating the use of Delphi techniques in areas of partial information. Insight was gained into group information processes. A meaningful estimate of the accuracy of a group response to a given question can be obtained by combining individual self-rating of competence on that question into a group rating. Adding this result to an observed relationship between accuracy and standard deviation makes it possible to attach accuracy scores to the products of a Delphi exercise. (R)

DALKEY, N. C.
Quality of Life
The Rand Corporation, P-3805, March 1968.

DALKEY, N. C.
Predicting the Future
The Rand Corporation, P-3948, October 1968.

Opinion is basic to long-range developmental forecasting. The difficulties (such as the influence of dominant individuals, noise, and group pressure for conformity) of obtaining a group opinion through traditional face-to-face interaction led to the development of the Delphi procedures, which are described in this paper. The characteristics of these procedures--anonymity, iteration with controlled feedback, and statistical group response--derive from Rand's discovery that simply the average of individual opinions, without discussion, tends to be more accurate than group opinion resulting from discussion. The experiments that led to these results involved almanac questions, such as, How many votes did Kennedy receive in the 1960 Presidential election in Texas? An initially wide range of answers were found to gradually converge, improving in accuracy in the majority of cases; the pattern of responses resembled a log-normal curve. Further studies will attempt to dampen the effect of group pressure while amplifying accuracy. (R)

DALKEY, N. C.
Experiments in Group Prediction
The Rand Corporation, P-3820, March 1968.

The use of the Delphi method for group prediction and estimating in a series of Rand experiments. The method has three basic features: (1) It elicits individual opinion, usually by questionnaire, but opinions are not attributed to specific individuals when communicated to the group. (2) It provides controlled feedback: An exercise is conducted in several rounds, opinions generated during one round being fed back to the group on the next round, usually in the form of statistical summaries. (3) Group opinion is expressed in terms of a statistical score. In most cases, there is a pronounced convergence of opinion with iteration; a wide spread on the initial round decreases monotonically on succeeding rounds, principally between the first and second. Where accuracy of response can be checked, it is shown to increase with iteration. Recent Rand experiments have focused on the use of information that can be readily verified as a means of further investigating the efficacy of the Delphi technique. (R)

DALKEY, N. C.
Delphi
The Rand Corporation, P-3704, October 1967. (See also P-3558.)

An outline of the Delphi technique of long-range forecasting by separately eliciting and refining the opinions of a group of advisors without contact among them, and calculating a statistical "group response." The procedure was designed to overcome the disadvantages common to committees and small groups. The experts reply to written questionnaires or an online computer, receive statistical feedback through formal lines of communication, and resubmit their estimates. Where the response is a number (such as a date or amount), the most useful index has been the median of the individual estimates. During the process, opinions do converge; where answers can be checked against reality, it is found that the median response tends to move in the direction of the true answer. Self-confidence is not correlated with individual performance, but the subgroup with the highest self-ratings for competence will consistently perform slightly better than the group as a whole. (R)

DALKEY, N. C., and B. BROWN
Comparison of Group Judgment Techniques with Short-Range Predictions and Almanac Questions
The Rand Corporation, R-678-ARPA, May 1971.

An experiment designed to discover whether the results of laboratory studies dealing with general (almanac) information are relevant to the applied case when the true answer is unknown. Using short-range prediction questions as subject matter, the experiment indicates that, in general, Delphi procedures are at least as effective with short-range prediction as they have been for almanac material. Eight groups, of about 20 each, of upper-classmen and college graduates were given short-range prediction questions to answer in a two-round Delphi exercise. Satisfactory answers were obtained for 32 of the 40 questions. Correlations between standard deviation and accuracy, and between group self-rating and accuracy, were significantly higher for the prediction than

for the almanac questions. Half the groups generated estimates of the three quartiles of the distribution; the other half generated point estimates. No significant difference was observed between these two kinds of estimates. (R)

DALKEY, N. C., B. BROWN, and S. W. COCHRAN
The Delphi Method, IV: Effect of Percentile Feedback and Feed-In of Relevant Facts
The Rand Corporation, RM-6118-PR, March 1970.

An investigation of the effect on group accuracy of two variations in the Delphi procedures. In these exercises, twenty general information questions are answered by two groups of respondents, who, after receiving some form of feedback, may revise their answers. In the first variation, feeding back individual percentiles resulted in no improvement over feeding back the median and quartiles of the group response. On the other hand, in the second variation, adding a relevant fact to the median and quartiles information resulted in a statistically significant increase in numerical accuracy. The number of changed answers was also greater, suggesting that introduction of a relevant fact strengthens motivation for revision. For a number of military concerns, such as long-range technological development assessment or future threat evaluation, the expertise of a group of decisionmakers is relied on. The Delphi studies are an effort to improve such judgments through refined procedures. (R)

DALKEY, N. C., B. BROWN, and S. W. COCHRAN
The Delphi Method III: Use of Self-Ratings to Improve Group Estimates
The Rand Corporation, RM-6115-PR, November 1969.

An analysis of the validity of using self-ratings as a technique for selecting more accurate subgroups in applications of the Delphi procedures for eliciting group judgments. A series of experiments was conducted using sixteen groups of upper-class and graduate college students answering almanac-type questions (twenty subjects per group and twenty questions per subject). The findings indicate that if the difference in average self-rating between the subgroups is substantial, and if the subgroups are held to reasonable size, both the degree of improvement and the total number of improvements are greater than when feedback alone is used. This study augments the results reported in RM-5888 and RM-5957. (R)

DALKEY, N. C., and O. HELMER
An Experimental Application of the Delphi Method to the Use of Experts
The Rand Corporation, RM-727-PR (Abridged), July 1962. (Also published in *Management Science*, Vol. 9, No. 3, April 1963, pp. 458-467.)

An abridgment and revision of RM-727, *The Use of Experts for the Estimation of Bombing Requirements: A Project Delphi Experiment.*

DALKEY, N. C., R. J. LEWIS, and D. SNYDER
Measurement and Analysis of the Quality of Life: With Exploratory Illustrations of Applications to Career and Transportation Choices
The Rand Corporation, RM-6228-DOT, August 1970.

The future of transportation and its impact on regional environment depends in part on the interaction with other programs such as housing, education, health and welfare. This memorandum summarizes attempts thus far to determine whether it is feasible to identify and measure underlying dimensions of the quality of life (QOL) and relate them to transportation findings. A model of individual QOL includes a set of general qualities of the stream of events occurring to an individual that largely determine his sense of well-being. Several group judgment (Delphi) studies produced relatively well-defined lists of such qualities, including self-respect, affection, security, health, achievement, novelty, freedom, comfort and aggression. Other studies compared lists of qualities with employment-environment opportunities and transportation choices. A number of approaches are suggested for future investigation, for example, in-depth time-event studies with small groups and cross-sectional national surveys based on a QOL model. (R)

DALKEY, N. C., and D. L. ROURKE
Experimental Assessment of Delphi Procedures with Group Value Judgments
The Rand Corporation, R-612-ARPA, February 1971. (See also RM-5888, RM-5957, RM-6118.)

One of a series of studies using Delphi procedures to aid decision-makers in dealing with value judgments. Previous studies have not clearly shown that there is an appropriate population of factual questions to compare with value judgments; the variability of performance on factual questions in large, depending on the type of questions asked. With this in mind, some comparisons were made: Two groups of UCLA students were asked to generate and rate lists of value categories that they considered important to higher education and the quality of life. Analyses showed that (1) distributions were generally single-peaked and roughly bell-shaped, (2) the correlations between different groups and different rating methods were high, and (3) the number of changes and degree of convergence for value judgments (reduction in standard deviation) were comparable to similar indices for factual judgments. The experiment supported the conclusion that Delphi procedures are appropriate for processing value material as well as factual material. (R)

DALKEY, N. C., D. L. ROURKE, R. LEWIS, and D. SNYDER
Studies in the Quality of Life: Delphi and Decision Making
D. D. Heath, Lexington, Mass., 1972.

DARIAN, JEAN-CLAUDE, and FRANCOISE MORIZE
"Delphi in the Assessment of Research and Development Projects"
Futures, October 1973, pp. 469-483.

DAVIS, RICHARD C.
"Organizing and Conducting Technological Forecasting in a Consumer Goods Firm"
In James R. Bright and M. E. F. Schoeman (eds.), *A Guide to Practical Technological Forecasting*, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1973.

The use of technological forecasting techniques for a consumer goods firm's product planning procedures is described. In addition to

describing Whirlpool's information storage and retrieval system, the author illustrates how Delphi has been used as an information gathering technique. Part of a sample questionnaire is illustrated. (B)

DAY, LAWRENCE H.
"The Future of Computer and Communications Services"
National Computer Conference and Exposition, New York, June 1973.

The future applications of computer and communications capabilities in the three application areas are described (education, business-travel communications tradeoffs, and the home). The projections discussed have been drawn from a number of Delphi studies--Bell Canada, Parsons & Williams, I.F.F. work, EDUCOM. Comparisons between the various forecasts are discussed. (B)

DAY, LAWRENCE H.
"Long Term Planning in Bell Canada"
Journal of Long Range Planning, London, 1973. (Submitted for publication.)

This paper describes the Business Planning process at Bell Canada. The use of many technological forecasting techniques in gathering input data is described. The use of Delphi as a part of this gathering of "futures" information is discussed. Sample Delphi questionnaires and results are illustrated. (B)

DAY, LAWRENCE H.
"Delphi Research in the Corporate Environment"
Delphi: The Bell Canada Experience, Bell Canada, October 1972.

This article examines the use of the Delphi technique in the corporate environment. The paper indicates that a broad range of Delphi studies have been conducted or sponsored by various corporations in North America and Europe. These studies were conducted by consultants or planning groups in the business firms. Four Delphi studies conducted by Bell Canada's Business Planning Group are examined. Results from these studies are illustrated. Several important issues that must be considered when conducting these studies in the corporate environment are considered. These include: (a) should corporations pay for this type of basic research? (b) how can the results from Delphi studies be best utilized in business? (c) misusing Delphi results in business, (d) in-house versus consultant conducted studies, and (e) the proprietary nature of business Delphi study results. The paper concludes with some projections on the future of Delphi in the corporate environment.

DEAN, B. V., and S. MATHIS
Analysis of the Exploratory Development Project Evaluation Experiment
Department of Operations Research, Case Western Reserve University, Technical Memorandum No. 165, 1969.

This report describes a study done for the Army Materiel Command using modified Delphi procedure to evaluate research and development projects, whose value is of course often of a very subjective nature. Data were collected, using a modified Delphi, from a twelve member panel and linear regression models of project value were constructed

as a function of eight critical factors. There is a description of the procedures used and conclusions drawn. Delphi has its usual attendant advantages, disadvantages and questions; the addition of regression analysis aids in the evaluation of results. (P)

DE BRIGARD, R., and O. HELMER
Some Potential Societal Developments: 1970-2000
Institute for the Future, R-7, April 1970.

DELPHI PANEL ON THE FUTURE OF LEISURE AND RECREATION, SET INC.
Los Angeles, 1972. (Multiclient Proprietary Study)

This study was conducted by SET, Inc. (Social Engineering Technology) for a group of clients interested in market opportunities resulting from increased leisure. (B)

DERIAN, JEAN-CLAUDE, and FRANCOISE MORIZE
"Delphi in the Assessment of Research and Development Projects"
Futures, October 1973, pp. 469-483.

DICKSON, PAUL
Think Tanks
Atheneum, New York, 1971.

This journalistic history and study of the growth of "think tanks" in the U.S. is mainly concerned with developments in the governmental field. The discussion of Delphi forecasting (pp. 313-18 and pp. 324-36) does mention that: "Delphi is now coming into widespread use in industry to determine new markets, possible new products for the future, and pitfalls to development." No detailed examples are given except for a reference to the TRW work. (B)

DOLE. S. H., et al
Establishment of a Long-Range Planning Capability
The Rand Corporation, RM-6151-NASA, September 1969.

An examination of some of the major problems of performing an effective long-range planning function within NASA and a survey of some of the techniques of systems analysis that might aid in the task of overall agency planning. Long-range objectives and policies are defined and developed, and the consequences of future decisions analyzed, by structuring the planning process into five procedural phases: input, projection, creative, analytical, and output. Concurrent supporting analysis is used to develop an information base for decisionmaking on alternative strategies. In this context, the major and some minor techniques of modern systematic analysis are surveyed to determine their applicability. Those clearly applicable are: (1) many classical systems analysis methods, and methods for coping with uncertainty; (2) Delphi procedures; (3) worth assessment techniques; (4) relevance trees and morphological analysis; (5) other forecasting techniques. The remaining approaches are either uncertain or clearly inapplicable. A long-range planning function would significantly aid the overall NASA program, but it would require continuing support by top management and coordination with related planning areas. (R)

DOLE, S. H., et al.
Methodologies for Analyzing the Comparative Effectiveness and Costs of Alternative Space Plans: Volume 1 (Summary Volume)
The Rand Corporation, RM-5656-NASA, August 1968. (Limited distribution.)

DOLE, S. H., et al.
Methodologies for Analyzing the Comparative Effectiveness and Costs of Alternative Space Plans: Volume 2 (Technical Volume)
The Rand Corporation, RM-5656-NASA, August 1968. (Limited distribution.)

DOYLE, FRANK J., and DANIEL Z. GOODWILL
"An Exploration of the Future in Educational Technology" in H. A. Stevenson, R. M. Stamp, and J. D. Wilson, (eds.), *The Best of Times The Worst of Times--Contemporary Issues in Canadian Education* Holt, Rinehart, and Winston, Montreal, 1972.

The material from this article is extracted from the Bell Canada Educational Delphi. Areas covered include changing values of society, general trends in terminal use, and the acceptance of computerized library systems, computer aides instruction systems, and visual display systems. (B)

DOYLE, FRANK J., and DANIEL Z. GOODWILL
An Exploration of the Future in Educational Technology
Bell Canada Business Planning, Montreal, January 1971. (Proprietary)

This study explores future acceptance of a number of visual and computer communications services: Computer assisted instruction, computerized library systems, and audio visual display systems. Other areas explored include value trends (1970-2000), chemical learning, evolution in school design, changing role of the teacher, and technology in the home. (B)

DOYLE, FRANK J., and DANIEL Z. GOODWILL
An Exploration of the Future in Medical Technology
Bell Canada Business Planning, Montreal, March 1971. (Proprietary)

This study explores future acceptance of services such as multiphasic screening. Computer assisted diagnosis, computerized medical libraries and terminal capabilities in medicine. Other areas explored included value trends, remote physiological monitoring, future technology in the home, and changing roles in the medical profession. (B)

DROR, Y.
"La Prediccion de lo Politicamente Posible"
Revista Espanola de la Opinion Publica, July - December 1970, pp. 21-22, 89-98. (See also The Rand Corporation, P-4044, April 1969.)

A reprint from *Futures*, issue not specified. The application of the Delphi method to factual political prognosis is considered. Factual political forecasting, contrary to politically oriented predictions, involves reference to an agent, an alternative policy and a political area. It can be approached via the following variables: principal agents, their capacities and intentions; actual and potential forces within the political area; the interactions between agent and increased political leverage; the critical influence of the masses. A three-fold theoretical schema based on the Delphi method is presented for political

forecasting. For details on the Delphi method, see O. Helmer, *Social Technology*, New York, New York, Basic Books, 1966. The three parts of the schema are: (a) direct estimates of political practicability, (b) conditions of political practicability and (c) estimates of the variables of political practicability. This method allows for evaluating the strength of a political possibility concerning each political alternative. A side benefit is the education of experts in prediction in general and in political analysis in particular. A word of warning is given concerning the capacity of unpredictable human forces to overcome apparently unsurmountable barriers. This introduces an element of the provisional into all political forecasting.

ENZER, SELWYN
Some Development in Plastics and Competing Materials by 1985
Institute for the Future, R-17, January 1971.

Owens-Corning sponsored this study which attempted "to focus upon possible combinations of material property changes that are likely to affect widespread material usage." The materials considered by the Delphi panel included: engineering plastics, general purpose and specialty plastics, glass fiber reinforced plastics, foamed plastics, and nonplastics. The study includes forecasts of U.S. plastic production, anticipated changes in properties of existing materials and developments in other important materials to 1985. (B)

ENZER, SELWYN
Some Prospects for Residential Housing by 1985
Institute for the Future, R-13, January 1971.

Owens-Corning Fiberglas sponsored this study to determine the most probable trends in residential housing in the coming 15 years and developments, actions, and policies that could alter probable trends. The panel results include forecasts of housing supply and demand, housing costs, and institutions and monetary aspects of housing, housing tenure, building codes, technological developments affecting residential housing and society. Panel charts and comments are included. (B)

ENZER, SELWYN
Delphi and Cross-Impact Techniques: An Effective Combination for Systematic Futures Analysis
Institute for the Future, WP-8, June 1970.

ENZER, SELWYN
A Case Study Using Forecasting as a Decisionmaking Aid
Institute for the Future, WP-2, December 1969.

ENZER, SELWYN, and R. DE BRIGARD
Issues and Opportunities in the State of Connecticut: 1970-2000
Institute for the Future, R-8, March 1970.

"The Exploration of the Future"
Réalités, No. 245, June 1966, pp. 50-58. Translated from the French by R. Neiswender, The Rand Corporation, P-3540, February 1967.

Translation from the French of an article in *Réalités* on the world of the future as visualized by participants in an international conference organized by *Réalités*. The article reports the work of a group of specialists whose interests are focused on a systematic exploration of the future, their opinions on the advantages and disadvantages of life in the 21st century, and their varying approaches to predicting and planning for the future. Among the forecasting methods described is the Delphi technique. In the opinion of *Réalités*, the time has come to establish both national and international institutes for prediction and planning. (R)

ENZER, SELWYN, DENNIS LITTLE, and F. D. LAZAR
Some Prospects for Social Change by 1985 and Their Impact on Time/Money Budgets
Institute for the Future, R-25, March 1972.

General Telephone and Electronics sponsored this study. The report outlines potential developments and trends likely to produce major changes in patterns of time/money expenditures in the next fifteen years. The study also had a secondary objective of experimenting with social forecasters using a modified Delphi approach. (B)

ESSO (EXXON)

Participants in a study and users of Delphi output in multi-industry study described by GLAZIER, et al. (B)

FARQUHAR, J. A.
A Preliminary Inquiry into the Software Estimation Process
The Rand Corporation, RM-6271-PR, August 1970.

Reviews the literature of software estimation and reports a small experiment comparing Delphi with face-to-face group judgment to predict the time necessary to program an information system--in this case, the Air Forces's PDSO (Personnel Data System--Officers). Planning software production is necessary but almost impossible at present. Cost to completion depends on many factors, some unknown at the time and all hard to quantify (the difficulty of the task, the programmer's ability and familiarity with the procedures involved, the degree of definition provided him, and about 80 other factors). The experiments undertaken failed to establish the utility of either estimation method. Primary recommendations for further research are: (1) more effective data collection, (2) analysis of characteristics of good estimators, and (3) formal inquiry into the techniques used by estimators. (R)

FELDMAN, N. E.
"Communication Satellite Output Devices"
The Microwave Journal, Nov. and Dec. 1965 issues.

FELDMAN, PHILIP
"Internal and External Delphi Panel Comparison"
Delphi: The Bell Canada Experience, Bell Canada, October 1972.

To determine whether there are significant differences in the responses of internal and external Delphi panels to identical questionnaires. To determine whether it will be necessary to go outside the company for panelists on future Delphi studies.

FELDMAN, PHILIP
A Technology Assessment of Computer-Assisted Instruction
Bell Canada, Business Planning, August 1972 (internal document).

This study has two main objectives: (a) to illustrate a methodology that will be of use for technology assessments, (b) to have a pilot assessment of the societal impacts of computer assisted instruction in post secondary institutions. Feldman uses input from the Bell Canada Educational and Business Delphi studies in the assessment. The report illustrates how Delphi study results can be used as means towards ends rather than ends in themselves. (B)

FISHBEIN, M. (Ed.)
Readings in Attitude Theory and Measurement, Wiley, New York, 1967.

"Forecasters Turn to Group Guesswork: Delphi Technique is Catching on with Corporations"
Business Week, No. 2115, March 1970, pp. 130-34.

This article describes recent expansion of the Rand, Institute for the Future, and National Industrial Conference Board activities. Corporate activities described include McDonald Douglas, Weyerhaeuser Co., Smith, Kline and French, and TRW. Most references are brief and the article is mainly survey in nature. The research of Dalkey at Rand is also discussed. (B)

GENERAL DYNAMICS, CONVAIR DIVISION

Discussed in Martino *Futurist* (1972) article. Three studies are mentioned. The Aerospace group organized a Delphi study on the future of fluidics to guide Convair R&D in the field. The second study examined the relative values of possible non-destructive testing techniques. This data was used for corporate decisions on the value of composite materials in aircraft construction. The third study produced a forecast of laser developments. (B)

GENERAL TELEPHONE AND ELECTRONICS CORPORATION
Sponsor of the Institute for the Future study: *Some Prospects for Social Change by 1985 and Their Impact on Time/Money Budget*, R-25, (Enzer, Little, and Luzer).

GLAZIER, FREDERICK, P., et al.
"A Multi-Industry Experience in Forecasting Optimum Technological Approach"
in James R. Bright and M.E.F. Schoeman (eds.) *A Guide to Practical Technological Forecasting*, Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1973.

This paper illustrates how the author used Delphi to develop a multi-industry, multi-firm approach to future technological requirements (Esso and Phillips Petroleum). Detailed examples of questionnaires and study results are illustrated in the paper. (B)

GOODMAN, JOEL M.
"Delphi and the Law of Diminishing Returns"
Technological Forecasting and Social Change 2, 1970, p. 225.

The purpose of this brief paper is to outline the author's thesis on the relationship between the number of Delphi studies conducted in an area and the amount of new material obtained by conducting further studies. In this brief, he mentions that Lockheed Aircraft Corp. conducted a Delphi study "for the purpose of identifying evolutionary and revolutionary technological concepts that might be anticipated during the remainder of the twentieth century." No results of the study are shown. (B)

GOODWILL, DANIEL Z.
"A Look at the Future Impact of Computer-Communications on Everyday Life"
Delphi: The Bell Canada Experience, Bell Canada, October 1972.

The computer communication revolution will have an impact on every aspect of our lives. Our concepts of work, work locatio.., home, leisure will undergo some rather startling changes over the next thirty years. This is the view of a group of individuals who participated in a recent Bell Canada Delphi study. It is very important for leaders in all sectors of the economy to be aware of the implications of these changes and to begin thinking about how we can implement these developments in an orderly and effective manner.

GOODWILL, DANIEL Z.
An Exploration of the Future Business Information Processing Technology
Bell Canada Business Planning, October 1971. (Proprietary)

This Delphi first explores the future of societal values and expected changes in business procedures. Technological services examined include management information services, mini and small computers, data processing, terminal capabilities, computer and communications services in the home, and the impact of technology on work locations. (B)

GORDON, T. J.
A Study of Potential Changes in Employee Benefits, Volume I: Summary and Conclusions
Institute for the Future, R-1, April 1969.

GORDON, T. J.
A Study of Potential Changes in Employee Benefits, Volume II: National and International Patterns
Institute for the Future, R-2, April 1969.

GORDON, T. J.
A Study of Potential Changes in Employee Benefits, Volume III: Delphi Study
Institute for the Future, R-3, April 1969.

GORDON. T. J.
A Study of Potential Changes in Employee Benefits, Volume IV: Appendices to the Delphi Study
Institute for the Future, R-4, April 1969, out of print.

GORDON, T. J., and R. H. AMENT
Forecasts of Some Technological and Scientific Developments and Their Societal Consequences
Institute for the Future, R-6, September 1969.

GORDON, T. J., and O. HELMER
Report on a Long-Range Forecasting Study
The Rand Corporation, P-2982, September 1964. (Also published as appendix to O. Helmer, *Social Technology*, Basic Books, New York, 1966.)
Description of an experimental trend-predicting exercise covering a time period as far as fifty years into the future. The Delphi technique is used in soliciting the opinions of experts in six areas: scientific breakthroughs, population growth, automation, space progress, probability and prevention of war, future weapon systems. Possible objections to the approach are also discussed. (R)

GOULD, JULIUS, and WILLIAM L. KOLB (eds.)
A Dictionary of the Social Sciences
Free Press of Glencoe, New York, 1964.

GRABBE, E. M., and D. L. PYKE
"An Evaluation of the Forecasting of Information Processing Technology and Applications"
Journal of Technological Forecasting and Social Change, (to be published)

HALL, P. D.
"Technological Forecasting for Computer Systems" in G. S. C. Wills et al., (eds.), *Technological Forecasting and Corporate Strategy*, Bradford University Press, 1969.
Discusses technological forecasting and Delphi. Several charts showing the computer Delphi for ICL are shown as well. (B)

HALL, T. W.
"Implementation of an Interactive Conference System"
AFIPS Conference Proceedings, Vol. 38, 1971 Spring Joint Computer Conference held from May 18-20, 1971 in Atlantic City, New Jersey, Montvale, New Jersey, AFIPS Press, 1971, pp. 217-229.

HAYDEN, SPENCER
"How Industry is Using Technological Forecasting"
Management Review, May 1970, pp. 4-15.
This article reports on the author's survey of 65 "progressive" (no definition) companies and their experiences with technological forecasting. Thirty methods were cited by the companies. While panel consensus methods (consulting groups of experts in person or writing) were found in 69 percent of the companies, formal Delphi was rated relatively unknown. Hayden found that 26 percent of the firms had used Delphi and that 71 percent of these firms had proved it useful. There are not any specific company references or sample Delphi results in the paper. (B)

HAYDON, B. W.
The Year 2000
The Rand Corporation, P-3571, March 1967.

A survey of the possible future of the year 2000 and the problems that advancing technology pose for civilized man. Rand's Delphi technique and the implications of some of the predictions obtained by the method are discussed: over-population and food production; nuclear power as a source of energy; air pollution; weather control; automation, education, and the home. Two compelling reasons exist for looking into the future: to detect danger signals so that action can be taken to forestall unpleasant events or conditions, and to avoid making mistakes. Decisions made today directly affect the future, and rational man must view the future as subject to his control. (R)

HELLER, F. A.
"Group Feedback Analysis: A Method of Field Research"
Psychological Bulletin, Vol. 72, No. 2, 1969, pp. 108-117.

"Group feedback analysis" combines a number of features from nomothetic and idiographic methods, thus overcoming some of the limitations of the mailed questionnaire. Certain advantages were described by reference to a number of research problems. Group feedback analysis has three interrelated stages: (1) individual research instruments are administered to a group of subjects; (2) some, or all, of the results from Stage 1 are fed back to the group as means and deviations; and (3) a discussion based on feedback of results is stimulated, recorded, and later content analyzed. Moving from unstructured to focused questioning, the investigator obtains a varied amount of information which acts both as a check and extension of the results obtained from Stage 1.

HELMER, O., and H. HELMER
Future Opportunities for Foundation Support
Institute for the Future, R-11, June 1970.

HELMER, O.
Long-Range Forecasting--Roles and Methods
Institute for the Future, P-7, May 1970.

HELMER, O.
Systematic Use of Expert Opinions
The Rand Corporation, P-3821, November 1967.

An explanation of the experimental Delphi technique, a systematic procedure for obtaining the opinions of experts on a particular subject. Four sets of questionnaires are used, each asking for successive refinements in the estimated answer to a given question. The interquartile range of the response is crucial. Respondents outside this range are invited to defend or reevaluate their answers, using the information feedback available. Refinements include subsidiary questions, attribution of differential weights to opinions, and the removal of systematic bias. Future applications may use automatic processing for opinions of panel members geographically remote from each other. Simplified versions of the Delphi technique can be used in face-to-face discussion; more complex versions tap the panelists' intuitive knowledge through heirarchical sets of expert opinions. (R)

HELMER, O.
Prospects of Technological Progress
The Rand Corporation, P-3643, 1967.

Discusses the role that the forecasting of technological developments plays in shaping the future of our society. Gives some predictions for the year 2000. Forecasting is an aid to decisionmakers, and can effect the future. Analysis of the future requires: (1) a survey of alternatives; (2) an analysis of preferences; (3) constructive policy research. He feels the prospects of socio-technological progress during the next third of this century are very high, due to increased scientific manpower, the computer and a reorientation toward policy-related research. Advent of social technology. Maybe a comprehensive theory of organizations.

"... we may well look forward to the emergence of a new breed of modern-day constructive utopians." (P)

HELMER, O.
Methodology of Societal Studies
The Rand Corporation, P-3611, June 1967.

The qualitative improvement of societal studies depends upon all-out acceptance of operations research techniques, adoption of a systems approach as a basic principle, a real effort toward interdisciplinary collaboration, and deliberate, intensive orientation toward the future. Competency in the techniques associated with the new methodology will be required. (R)

HELMER, O.
The Future of Science
The Rand Corporation, P-3607, May 1967.

The application of the techniques of the physical sciences to the social sciences, coupled with the increasing capability and refinement of the computer, is presented as a possible method of solving socio-political problems in the future. The increasingly symbiotic relationship between man and machine and the refinement of interdisciplinary, operations-analytical techniques in the social sciences, particularly the Delphi technique and game theory will produce breakthroughs and new forms of procedure in the scientific establishment. (R)

HELMER, O.
New Developments in Early Forecasting of Public Problems: A New Intellectual Climate
The Rand Corporation, P-3576, April 1967. (Also in *Vital Speeches*, Vol. 33, 1967, pp. 497-499.)

A report of philosophical, pragmatic, and methodological changes in world attitude toward the future--all favoring positive long-range planning. The second computer revolution is leading to true man-machine symbiosis. The social sciences are turning to an interdisciplinary systems approach to the solution of socio-political problems, using mathematical models, simulation procedures, and a systematic approach to the utilization of expert opinions. To arrive at a positive payoff for all requires reasonable expectation that cooperative moves will meet with a cooperative response. The revolutionary reorientation in

the social sciences gives hope that in the next generation this attitude will extend into the uncivilized area of international relations. (Prepared for presentation at a Public Affairs Perspective Conference under National Industrial Conference Board auspices, New York, April 1967.) (R)

HELMER, O.
Analysis of the Future: The Delphi Method
The Rand Corporation, P-3558, March 1967.

A description of the Delphi technique which attempts to make effective use of informed intuitive judgment in long-range forecasting. The Delphi method in its simplest form solicits the opinions of experts through a series of carefully designed questionnaires interspersed with information and opinion feedback. A convergence of opinion has been observed in the majority of cases where the Delphi approach has been used. In a few of the cases where no convergence toward a relatively narrow interval of values took place, opinions began to polarize around two distinct values; two schools of thought regarding a particular issue seemed to emerge. Refinements that have been made in the Delphi technique consist of the introduction of weighted opinions and use of the technique in conjunction with a simulated decisionmaking process. (R)

HELMER, O.
A Use of Simulation for the Study of Future Values
The Rand Corporation, P-3443, September 1966.

HELMER, O.
The Use of the Delphi Technique in Problems of Educational Innovations
The Rand Corporation, P-3499, December 1966.

A description of the Delphi technique, a method for the systematic solicitation and collation of expert opinions, and its applications to educational planning. Delphi pilot experiments carried out in an Educational Innovations Seminar, UCLA, apply the Delphi technique of long-range forecasting to proposals for innovations in educational methods and budget allocations to achieve these innovations. (R)

HELMER, O.
Social Technology
Basic Books, New York, 1966. (See also The Rand Corporation, P-3063, February 1965.

A reappraisal of methodology in the social sciences with specific proposals for modifications of traditional procedures. The paper suggests that social scientists explore the possibilities of operations research approaches, of operational model building, and of expert opinion. It also suggests procedures in the areas of urban renewal, educational reform, political, and long-range economic forecasting. (R)

HELMER, O.
Convergence of Expert Consensus Through Feedback
The Rand Corporation, P-2973, September 1964.

A discussion of studies directed toward the improved use of expert opinions in operations research. The Delphi method as applied to consensus research is discussed and the results of experiments directed toward convergence of expert opinions presented. (R)

HELMER, O.
The Systematic Use of Expert Judgment in Operations Research
The Rand Corporation, P-2795, September 1963.

A consideration of operations research as a science, though an inexact one. The operations analyst, as opposed to the pure scientist, emphasizes control rather than understanding. Judgment in constructing and applying operations-analytical models should be as expert and its applications as systematic as possible. Improved methods of identifying and measuring expertness and of employing experts efficiently are needed. In particular, the use of groups of experts by consensus techniques, by the Delphi technique, and by simulation procedures and operational gaming should be further refined. (R)

HELMER, O.
The Outcome of a Recent Experiment in Prediction
The Rand Corporation, 1952. (Unpublished)

HELMER, O, T. J. GORDON, S. ENZER, R. DE BRIGARD, and R. ROCHBERG
Development of Long-Range Forecasting Methods for Connecticut: A Summary
Institute for the Future, R-5, September 1969.

HELMER, O., and E. S. QUADE
An Approach to the Study of a Developing Economy by Operational Gaming
The Rand Corporation, P-2718, 1963.

This is one of the earlier papers. It "... considers the possible use of operational gaming, or simulation using human players, to examine an economy as a whole." "[Operational gaming] undoubtedly is most useful when applied with a clear objective in mind to well-structured problems based on abundant data." However, there is more chance in a developing economy of applying operations research to "strategic" problems. The model is seen essentially as a communications medium. Provides a good argument for use of simulation before one has an adequate theory. Discusses computer simulation, and the objections to it, then advantages of gaming, where use of computers is minimal, reliance on the intuitive expertise of specialists, and emphasis is on clearer problem formulation. For gaming, break down economy into sectors. Include intangibles, also get feedback.

They do not see any immediately usable output. (P)

HELMER, O., and N. H. RESCHER
On the Epistemology of the Inexact Sciences
The Rand Corporation, R-353, February 1960.

A new epistemological approach to the inexact sciences, which include applied physical sciences, such as engineering and medicine, as well as most of the social sciences. The purpose of all science is to explain past events and to predict future ones in an objective manner.

While explanation and prediction have the same logical structure in the exact sciences, this is not so in the inexact sciences. This fact leads to the development of specifically predictive instrumentalities in these fields and to various methodological innovations. Among these are the systematic employment of expert judgment and the use of pseudoexperimentation, involving simulation processes, and in particular, operational gaming. (R)

HERCULES POWER COMPANY
See references by PARKER and WILLS for elaboration.

HERSCH, CHARLES
"The Discontent Explosion in Mental Health"
American Psychologist, 1969, pp. 597-606.

HILLS, L. S.
"Delphi Technique: A Tool for Business Forecasting"
AACE Bulletin, Vol. 12, No. 2, 1970, p. 48.

HONEYWELL
See the Alderson and Sproull article describing the use of the PATTERN Technique at Honeywell. This technique is partly related to Delphi as it uses panels of experts to develop and relevance numbers. The McGlauchlin paper describes some Honeywell Delphi experience in more detail. (B)

ICL (INTERNATIONAL COMPUTERS LTD.)
See references and results summarized by WILLS and by HALL.

IFIP-INTERNATIONAL FEDERATION OF INFORMATION PROCESSING
See the study conducted by PARSONS and WILLIAMS--"Forecast 1968-2000 ..."

IKLÉ, F. C.
Social Forecasting and the Problem of Changing Values, with Special Reference to Soviet and East European Writings
The Rand Corporation, P-4550, January 1971.

JANTSCH, E.
Technological Forecasting in Perspective: A Framework for Technological Forecasting, Its Techniques, and Organization: A Description of Activities and Annotated Bibliography
Organization for Economic Cooperation and Development, Paris, 1967.

THE JOHNS HOPKINS UNIVERSITY
"Provide Prognostic Epidemiologic Estimates"
Department of Medical Care and Hospitals, 1970.

JOLSON, MARVIN A., and GERALD L. ROSSOW
"The Delphi Process in Marketing Decision Making"
Journal of Marketing Research Vol. 8, November 1971, pp. 443-448.

Describes the use of Delphi in an experiment conducted for the Pace Computing Corporation in estimating market demand for its services. Part of the article describes the authors' evaluation of the techniques (similar to Rand experiments). Concludes that the technique can be used for marketing purposes and as an internal communications tool in a business. (B)

KAPLAN, A., A. L. SKOGSTAD, and M. A. GIRSHICK
"The Prediction of Social and Technological Events"
Public Opinion Quarterly, Vol. 14, No. 1, 1950, pp. 93-110.
(See also The Rand Corporation, P-93, 1949.)

A group of individuals with a generally high education level was asked to make predictions concerning a large number of future events in order to investigate certain aspects of the use of expert opinion in policymaking. In particular, the following questions were explored: how good are expert predictions in areas germane to policy; how can such predictions be improved; and how can the reliability of a given opinion be appraised beforehand? While the design of the study makes projection of its results on other situations questionable, it throws considerable light on the problems involved. Among other results, it was found that confidence in prediction does not necessarily show a correlation with success in prediction, that predictions made by groups of people are more likely to be right than predictions made by the same individuals working alone, and that the reliability of predictions can be appraised to some extent by examining the character of the justifications given for them.

The success of the collective psychological methods, i.e., group discussion and iteration, was duplicated by using statistical methods such as averaging on the individual results. (P)

KIMBLE, R.
Delphic Forecasting of Critical Personnel Requirements
U.S. Army Electronics Command, Fort Monmouth, New Jersey, November 1969.

KNOCK, RICHART T.
Electro-Optical Technology
Long Range Planning Service Report 468, Stanford Research Institute, August 1972, (Client proprietary).

This report illustrates how Delphi material can be integrated with other data in a research report on a specific subject. The Delphi data in this report was obtained from the TRW PROBE study and from the Japanese Science and Technology Agency--"Report on Technological Goods; September 1971." The Delphi summary chart presents 43 forecasts on 13 major industrial applications of electro-optical technology (pages 8-9). (B)

KOCHMAN, A. F.
An Investigation of the Delphi Forecasting Technique with Emphasis Upon Educational Need Assessment
Master's Thesis, Chapman College, 1968.

KOPKIND, A.
"The Future-Planners"
The New Republic, February 1967, pp. 19-23.

LACHMANN, OLE
"Personnel Administration in 1980: A Delphi Study"
Long-Range Planning, Vol. 5, No. 2, Oxford, England, June 1972.
This article reviews two studies on trends in personnel management. This Danish study was first prepared as input to a conference. Sample results are shown. The second study was an extension of the first one but using employees from the printing firm CON-FORM. Results from the CON-FORM study are also shown in the paper. (B)

LANSDOWNE, Z. F.
Analysis of Intercity Transport Improvements: Forecasting Demand and Evaluation User Benefits,
The Rand Corporation, RM-6255, May 1970.

LIKERT, R.,
The Human Organization
McGraw Hill, New York, 1967.

LING-TEMCO-VAUGHT INC. (LTV)
Reference and sample questionnaire page shown in Wills Technological Forecasting. Martino also references this study in his book. The Bright book "A Brief Introduction..." uses the LTV Delphi extensively as a case example. Detailed forecasts and methodology is outlined. Cetron also uses LTV experience in his book. (B)

LING-TEMCO-VOUGHT
Continuing Use of the Delphi
Proprietary document.

LORGE, I. M., D. FOX, J. DAVITZ, and T. BRENNER
"Survey of Studies Contrasting the Quality of Group Performance and Individual Performance, 1920-1957"
Psychological Bulletin, Vol. 55, 1958, pp. 337-372.
Research contrasting the quality of group performance with individual performance in each of the following general topic areas has been examined in this paper: judgment, learning, social facilitation, problem solving, memory, size of group, problem solving in more realistic situations, and productivity. Recent theoretical and methodological considerations as well as discussions of group types are included. Research weaknesses and theoretical problems are discussed.

LOCKHEED AIRCRAFT CORPORATION
See reference in paper by GOODMAN.

LUDLOW, J. D.
Sea Grant Delphi Exercises: Techniques for Utilizing Informed Judgments of a Multi-Disciplinary Team of Researchers
Bureau of Business Research, University of Michigan, Working Paper 22, January 1971.

Application of Delphi to regional and urban planning.

MACMILLAN BLOEDEL LTD.
Sponsor of the Institute for the Future Study R-16: *The Future of Newsprint* (Paul Baran).

MAIER, N. R. F.
"Assets and Liabilities in Group Problem Solving: The Need for an Integrative Function"
Psychological Review, Vol. 74, No. 4, July 1967, pp. 239-249.
Research on group problem solving reveals that the group has both advantages and disadvantages over individual problem solving. If the potentials for group problem solving can be exploited and if its deficiencies can be avoided, it follows that group problem solving can attain a level of proficiency not ordinarily achieved. The requirement for achieving this level of group performance seems to hinge on developing a style of discussion leadership which maximizes the group's assets and minimizes its liabilities. Since members possess the essential ingredients for the solutions, the deficiencies that appear in group solutions reside in the processes by which group solutions develop. These processes can determine whether the group functions effectively or ineffectively. The critical factor in a group's potential is organization and integration. With training, a leader can supply these functions and serve as the group's central nervous system, thus permitting the group to emerge as a highly efficient entity. Extensive bibliography. (P)

MARIEN, M. (Ed.)
The Hot List Delphi: An Exploratory Survey of Essential Reading for the Future
Educational Policy Research Center, Syracuse University Research Corporation, 1972.
A report listing 236 books and articles, of which 192 were rated by a panel of fourteen futurists. "Recommended to anyone seeking guidance to the best books to read in futuristics."

MARTINO, J. P.
"How the Soviets Forecast Technology"
The Futurist, Vol. 7, No. 1, February 1973, pp. 30-31.

MARTINO, J. P.
Technological Forecasting for Decisionmaking
American Elsevier, New York, 1972.
This detailed treatment of the subject (750 pp.) examines all aspects of technological forecasting. The chapter on Delphi uses corporate studies as examples in outlining how Delphi studies have been conducted. Business studies shown include those of Smith, Kline, and French, TRW, Parsons and Williams, and LTV. Many of the specific analyses of the technique itself were based upon examining the raw TRW data provided to Martino. (B)

MARTINO, J. P.
"Technological Forecasting is Alive and Well in Industry"
The Futurist, August 1972.

In this article Martino discusses the activities of several companies who have maintained technological forecasting activities during the recent economic recession while other firms stopped these studies. Delphi studies at the Convair Division of General Dynamics are outlined. The Bell Canada use of Delphi study results for Business Planning purposes is also discussed. Martino concludes that the survival of technological forecasting in the corporate environment is dependent upon making the results useful for decisionmaking purposes. (B)

MARTINO, J. P.
"What Computers May Do Tomorrow"
The Futurist, October 1969, p. 134-135.

The author reviews the Parsons and Williams Delphi study. The first part of the article outlines the background and conduct of the study (panel size, number of rounds and events forecast, panel make-up, etc.). The second part reviews some of the study findinge. One chart is included. (B)

MARTINO, J. P.
An Experiment with the Delphi Procedure for Long-Range Forecasting
Office of Scientific Research, U.S. Air Force, AFSOR 670175, 1967.

MASLOW, ABRAHAM
The Farther Reaches of Human Nature: Towards A Psychology of Being
Viking Press, New York, 1971, Princeton, N. J., Van Nostrand, 1962, 1968, Second Edition.

MASON, R. O.
"A Dialectical Approach to Strategic Planning"
Management Science, Vol. 15, No. 8, pp. B-403-414, 1969.

MCDONNEL DOUGLAS, DOUGLAS AIRCRAFT DIVISION

Referenced in the *Business Week* article. This study forecast the future of commercial air transportation. Revenue forecast on passenger and cargo operations were developed. One purpose of their forecasts is to determine the date when all-cargo jumbo jets should be introduced. (B)

MCGLAUCHLIN, LAURENCE D.
"Technological Audits: An Aid to Research Planning" in James R. Bright and M. E. F. Schoeman (eds.), *A Guide to Practical Technological Forecasting*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1973.

A use of Delphi at Honeywell for corporate-wide technological forecasting efforts is described. This use of Delphi is part of an overall process of R&D planning at Honeywell. Questionnaire samples, the use and value of Delphi as an internal communications tool is also outlined. (B)

MCGREGOR, DOUGLAS
"The Major Determinants of the Prediction of Social Events"
Journal of Abnormal and Social Psychology, Vol. 33, 1938, pp. 179-204.

MCLOUGHLIN, W. G.
Product Cycle Planning
Presented at the Technological Forecasting Conference, Lake Placid, New York, 1969.

MILKOVICH, G. T., A. J. ANNONI, and T. A. MAHONEY
The Use of the Delphi Procedures in Manpower Forecasting
Minnesota University Minneapolis Center for the Study of Organizational Performance and Human Effectiveness, TR-7007 (AD747 651), 1972. Presented at American Meetings of the Institute of Management Sciences, October 12, 1971.

A case study was made in the development, implementation and evaluation of the Delphi technique, which systematically makes use of expert judgment in generating manpower forecasts. This case study was conducted in a large national retail organization on professional manpower. The results of the Delphi technique using an expert and a naive panel are compared with results generated by a conventional regression based model and the actual experience of the organization, which serves as the criterion. The study also analyzes the informational elements used by experts during the Delphi procedures and develops a model based on these elements. The usefulness of the Delphi in generating manpower forecasting models is discussed.

MITROFF, I. I.
"A Communication Model of Dialectical Inquiring Systems--A Strategy for Strategic Planning"
Management Sciences, Vol. 17, No. 10, June 1971, pp. B-634-648.

MONSANTO
CETRON references the Monsanto work.

MOORE, C. G., and H. P. POMREHN
Technical Forecasting of Marine Transportation Systems 1970-2000
University of Southern California, June 1969.

MORRIS, P. A.
Bayesian Expert Resolution
Doctoral Dissertation, Stanford University, 1971.

NANUS, BURT, LELAND M. WOOTEN, and HAROLD BORKO
The Social Implications of the Use of Computers Across National Boundaries
AFIPS Press, Montvale, New Jersey, 1973 (forthcoming).

In all, more than seventy areas of potential impact were explored with the panel. The panel strongly agreed that world economic development will be greatly affected by the globalization of information processing systems and that the use of computers across national boundaries in both the public and the private sectors will expand very greatly in

the next two decades. Among the many specific conclusions of the panel are the following:

1. New institutions will be required at the multinational level to resolve disputes over the transmission of data across national boundaries, to develop regulations concerning the activities of multinational data banks, to provide individual safeguards, and to deal with the problems of standardization of data transmission facilities and capabilities.

2. The use of multinational computer systems will tend to enhance the power of multinational corporations vis-a-vis the nation-state while at the same time contributing to a growing uniformity of business practices throughout the world.

3. Within the highly industrialized societies, many people will find themselves in some form of man-machine relationship, often involving multinational communications, within the next decade. These interactions may be for education, health, library, business or other reasons but the net effects will be the enhancement of shared beliefs and values and a growing sense of interdependence on matters of the most fundamental nature.

4. The use of multinational computer systems may tend to enhance the economic interests of the information rich, wealthier nations at the expense of the information poor, but in the long run, the use of such systems will increase the technological options available to the less developed countries and speed up their ability to industrialize.

NATIONAL RESEARCH COUNCIL, NATIONAL MATERIALS ADVISORY BOARD, COMMITTEE ON TECHNOLOGICAL FORECASTING
Technological Forecasting and Engineering Materials
NMAB-279, December 1970.

Comparison of Delphi with other planning tools.

NATIONAL TRAINING LABORATORIES
Group Development
(Bradford, Leland Powers), Washington National Training Laboratory, National Education Association, 1961.

NORTH, H. Q.
A Probe of TRW's Future: The Next 20 Years
TRW Systems, July 1969, (proprietary document).

NORTH, H. Q.
Delphi Forecasting in Electronics
Thompson Ramo, Wooldridge, 1968, unpublished.

NORTH, H. Q., and D. L. Pyke
"Probes of the Technological Future"
Harvard Business Review, Vol. 47, May 1969, pp. 68-82.

Implications of technological change. Step-through of the PROBE and PROBE 2 Delphi studies. Similar charts to IEEE article. Appendix on the basic Delphi method. (B)

NORTH, H. Q., and D. L. PYKE
"Technological Forecasting in Planning for Company Growth"
IEEE Spectrum, January 1969.

A review of the use of TRW's Delphi modification (PROBE). The planning system at TRW is also illustrated in several detailed flow charts. No detailed results of the studies are shown. (B)

NORTH, H. Q., and D. L. PYKE
"Technology, the Chicken--Corporate Goals, the Egg"
Technological Forecasting for Industry and Government, James R. Bright (ed.), Prentice-Hall, Englewood Cliffs, New Jersey, 1968.
Technology and goals are interrelated, rather than one causing the other. (P)

OFFICE OF HEALTH ECONOMICS
Medicines in the 1990's--A Technological Forecast
London, October 1969.

OVERBURY, R. E.
"Technological Forecasting: A Criticism of the Delphi Technique"
Long-Range Planning, Vol. 1, No. 4, June 1969, pp. 76-77.

OWENS-CORNING FIBERGLAS CORPORATION
Sponsor of two Institute for the Future studies: R-13, *Some Prospects for Residential Housing by 1985* (ENZER) and R-17, *Some Developments in Plastics and Competing Materials by 1985* (ENZER).

PACE COMPUTING CORPORATION
See article by JOLSON and ROSSOW.

PACKARD, KARLE S.
"Impact of an Emerging Technology on Company Operations"
A Guide to Practical Technological Forecasting, James R. Bright, and M. E. F. Schoeman, (eds.), Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1973.
The author describes the use of a Delphi study as part of a technological forecasting exercise at AIL. The forecasting exercise was concerned with the number of electronic circuits that could be included on LSI chips. Other techniques used were horizontal decision matrices, trend extrapolations, and substitution analysis. (P)

PARKER, E. F.
"British Chemical Industry in the 1980's--A Delphi Method Profile"
Chemistry and Industry, January 1970.
Continuation of Parker's description of the use of Delphi at the Hercules Powder Company. Detailed tables and charts showing study results are included. The results are also interpreted in the text. An appendix reproduces the Third Round Questionnaire. (B)

PARKER, E. F.
"Some Experience with the Application of the Delphi Method"
Chemistry and Industry, No. 38, London, September 1969, pp. 1317-1319.
This is a general review of Parker's experience with Delphi at the Hercules Powder Company. Some basic results of the study are also referenced. (B)

PARSONS, HENRY MCILVAINE
Man-Machine System Experiments
The John Hopkins Press, Baltimore, Maryland, 1972.

Man-machine system experiments comprise a field of research which has developed in the last two decades in response to new technologies such as radar and computers. Despite the magnitude and importance of this field, little about its methodology has appeared in print, nor have reports of the experiments themselves been generally accessible.

In contrast to other human factors studies, a large-scale man-machine system experiment typically investigates coordinated team performance that may involve a dozen or more system operators. Although the elaborate simulation required if often computer-based, in this research--unlike all-computer simulations--actual people operate actual machines. For some experiments major facilities have been created; for others, the system itself has been the laboratory.

This book describes more than 200 of these complex experiments and the forty-odd programs in which they occurred at universities, non-profit institutions, and government research agencies. The author has drawn on information from some 600 references, numerous investigators and consultants, and first-hand experience to probe what he once called "fifty million dollars of buried research."

He also analyzes systematically the methodological problems that result from experimentation on air defense systems, air traffic control systems, logistics organizations, space flight, battlefield operations, police dispatching, and even communications between heads of state. He includes experimental results concerning system procedures, training methods, team composition, organizational adaptation, complex decision-making, man-machine capabilities, and many aspects of design including degrees of computer automation.

This comprehensive survey can guide those who plan similar experiments in the future. It will interest psychologists and engineers engaged in human factors research and application, data processing and operations research specialists in system analysis and simulation, teachers of the experimental method, and all those concerned with the roles of man and machine in today's fast-developing technologies.

PARSONS and WILLIAMS
Forecast 1968-2000 of Computer Developments and Applications
Copenhagen, 1968.

This Delphi study was prepared by Parsons and Williams (Danish consultants) as an input to panel discussions at a conference sponsored by the International Federation of Information Processing Societies. A wide variety of computer applications and technological advances are explored. Study respondents were individuals planning to attend the conference. (B)

PHILLIPS PETROLEUM

Participants in a study and users of Delphi output in a multi-industry study described by GLAZIER et al.

PILL, JURI
"The Delphi Method: Substance, Contexts, A Critique and an Annotated Bibliography"
Socio-Economic Planning Sciences, Vol. 5, 1971, pp. 57-71.

While this paper only makes passing reference to industrial Delphi work (i.e., TRW), it does contain an excellent annotated bibliography on the technique itself. (B)

PWG PUBLICATIONS
"Japanese Futures Research Trend"
Technology Forecasts and Technology Surveys, June 1973, pp. 2-3.

PYKE, DONALD L.
"Mapping--A System Concept for Displaying Alternatives"
A Guide to Practical Technological Forecasting, James R. Bright and M. E. F. Schoeman, (eds.), Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1973.

Pyke elaborates further on the TRW experience with Delphi. The emphasis in this paper is on the use of the forecasts from the PROBE II. Pyke shows how the various forecasts can be considered in relationship to each other through a mapping technique (a PERT type chart). (B)

PYKE, DONALD L.
"Technological Forecasting: A Framework for Consideration"
Futures, Vol. 2, No. 4, December 1970, pp. 327-331.

PYKE, D. L.
"Practical Approaches to Delphi"
Futures, Vol. 2, No. 2, June 1970, pp. 143-152.

QUADE, E. S.
On the Limitations of Quantitative Analysis
The Rand Corporation, P-4530, December 1970.

Today we often hear that to meet the many challenges to our society we need only turn to systems analysis, operations research, "aerospace technology"--in other words, to supposedly objective quantitative analysis. Few realize that quantitative analysis ultimately reduces to an orderly series of judgments in defining the problem, selecting hypotheses and approaches, making assumptions, determining the "facts," assigning numerical values and relationships. Unfortunately, the cost/benefit criteria usually differ greatly from those used by political participants. Operational gaming with role-playing allows for the interplay of different viewpoints and for making decisions in context as the need arises. For situations still harder to model, expert judgment is commonly sought. Delphi questioning, featuring anonymity, iteration, controlled feedback, and statistical response, is increasingly used, but more experimental work is needed. Viewed as a method for investigating problems rather than solving them, analysis is nearly always useful. (R)

QUADE, E. S.
An Extended Concept of "Model"
The Rand Corporation, P-4427, July 1970.

A broad view of the traditional operations model, suggesting that any device be accepted as a model if it provides a logical means of predicting and comparing the outcomes of alternative actions, regardless of its representative features or its efficiency in optimization. Illustrative of this extended type of model is Delphi, an iterative procedure for eliciting and refining the opinions of a group of people by means of a series of questionnaires. While much remains to be learned about Delphi and the use of expertise, its potential is considerably wider than published applications indicate. Industrial and urban planners, research managers, and policymakers have shown extensive interest. Suggested applications range from the drafting of diplomatic notes and long-range political forecasting to determining what products to market. Because it can be used to allocate resources rationally and to force explicit thinking about the measurement of benefits, Delphi offers a hope of introducing cost-effectiveness thinking to a wide range of problems where conventional models are difficult to formulate. (R)

QUADE, E. S.
Cost-Effectiveness: Some Trends in Analysis
The Rand Corporation, P-3529-1, March 1970.

Four trends can be seen in the development of cost-effectiveness analysis. (1) Close man-computer interaction--through on-line, time-shared systems with individual consoles, natural language, graphic input, and disc-stored submodels--enables the user to modify his program instantly, even during execution. Hence, it facilitates the numerous parametric investigations and sensitivity analyses necessary in ranking alternatives. (2) The theory of n-person games accommodates the models with many insignificant players so common in cost-effectiveness analyses. Judgment and intuition guide both quantitative and nonquantitative aspects of an analysis, especially those with high social and political content. (3) Expertise can be systematically exploited by Delphi, an iterative procedure for eliciting and refining group opinions through a series of questionnaires with anonymous response. (4) Analysts are now facilitating implementation of their studies by investigating the potential effectiveness, political feasibility, the internal organizational acceptability of their recommendations, and the relationships between individual and organizational needs and objectives. (R)

QUINN, JAMES BRIAN
"Technological Forecasting"
Harvard Business Review Forecast Series, No. 21215 (1971), pp. 51-53.

REISMAN, A., S. J. MANTEL, JR., B. V. DEAN, and N. EISENBERG
Evaluation and Budgeting Model for a System of Social Agencies
Department of Operations Research, Case Western Reserve University, Technical Memorandum No. 167, 1969.

This report describes a relatively large-scale experiment on scaling group opinion using a procedure based essentially on Delphi, in that it used controlled feedback and anonymity. The main difference was that an overt attempt was made to achieve consensus, rather than stopping at a prescribed number of rounds.

The objective of the study was to estimate the relative values of services performed by the agencies of the Jewish Community Federation (JCF) of Cleveland, as perceived by the members of that community. There was a test run using members of the research staff who were themselves members of the community, then a procedure using the JCF staff, and finally a panel of community lay leaders. There was a ranking of 250 "client-service packages." Correlation was surprisingly good among the three groups. The experiment demonstrated the feasibility of obtaining a set of meaningful community standards, meaningful in the sense that the representatives of the community accepted them as valid and realistic. The process also proved to have value in itself as a communication device, and to make it easier to use the results, since the participants had helped generate them. (P)

REISMAN, A., and M. I. TAFT
Computer Time-Saving Applications in Economic Analysis: An Integrated Approach
Department of Operations Research, Case Western Reserve University, Technical Memorandum No. 151, 1969.

A generalized evaluation program known as EVAL, and a general "present worth" program known as CERBS have been developed in the FORTRAN IV Computer language and are currently operational on a time-sharing or batch basis. The programs lead an individual or a group of people through a systematic step-by-step procedure for evaluating the relative value (utility including present worth) of each of a given set of alternatives with respect to explicitly stated goals and objectives. Applications in such areas as equipment design, selection, optimization and replacement; personnel administration; and allocation of resources are presented. (P)

REISMAN, A., and M. I. TAFT
On a Computer-Aided System Approach to Personnel Administration
Department of Operations Research, Case Western Reserve University, Technical Memorandum No. 147, 1968.

A systematic evaluation methodology has been developed which integrates some major concepts from value, utility, decision, subjective probability theories and the Delphi method for obtaining a consensus of opinions. These theories are applied to the process of evaluation of personnel for recruitment, promotions, merit raises, transfer, salary administration, training and development. The model requires and utilizes as inputs explicitly stated sets of long-range goals, short-range objectives, resource needs, evaluative criteria, weighting and utility functions, as well as the subjective judgments of appropriate evaluators. The processing of this information may be implemented by manual calculations, batch processing on an IBM 1620 computer, or by direct simulation on a large time-sharing computer system. By utilizing standard statistical procedures and the decision rule to maximize expected utility, the methodology produces the type of output information required for rational decisionmaking. (P)

RESCHER, N.
Delphi and Values
The Rand Corporation, P-4182, September 1969.

An analysis of the utility of the Delphi method for determining group values. Delphi, a process for eliciting group opinion by a series of questionnaires with selective feedback of earlier responses, has traditionally been used for assessing expert opinion about factual questions. In the area of values, Delphi is more appropriate for exploring group *evaluation* (which involves rational criteria) than group value *judgment* (which depends on emotional response). Delphi can also be an instrument for (1) determining the values most relevant to the decision by focusing upon the reasons (and their relative weights) for making certain choices; (2) discovering areas of consensus within general value conflicts; (3) finding subgroups of variant opinion; (4) resolving conflicts of interest by questioning noninvolved third parties. However, the utility of Delphi for ascertaining values is limited by the impossibility of checking the group value judgments against "the actual facts." (R)

RESCHER, N.
The Future as an Object of Research
The Rand Corporation, P-3593, 1967

Limits "the perspective to two sectors of particular interest alike to makers of public policy and to reflective citizens in the modern world: science and technology on the one hand, and our human and social environment upon the other."

It is now in fashion to speculate about the future. The Futures Industry. The problem of predictive methodology. Some major difficulties for prediction: feedback, change occurrence, fashions, values, the problems of data.

This is an excellent summary of the future as an object of research. (P)

ROCHBERG, R.
Some Comments on the Problems of Self-Affecting Prediction
The Rand Corporation, P-3735, 1967.

This paper attempts to analyze and clarify the possible interaction of prediction and event. (1) Is the interaction of prediction and event predictable? (2) Can the interaction of prediction and event be controlled? (3) How and to what extent should people making predictions try to take into account the possible effects of their predictions? (4) Is the entire phenomenon of self-affecting predictions of any practical use?

He uses flow charts, a probabilistic model, a cybernetic model and discusses various types of prediction. In general, this is a good overview of the problem, without any real theory of answers. (P)

ROGERS, C.
Encounter Groups
First Edition, Harper and Row, New York, 1970.

ROGERS, C.
On Becoming a Person
Houghton Mifflin, Boston, 1961.

ROSOVE, PERRY E.
The Use of Contextual Mapping to Support Long-Range Educational Policy Making
SP-3026, System Development Corporation, Santa Monica, California, December 1967.

SACKMAN, H. and R. CITRENBAUM (Eds)
Online Planning
Prentice-Hall, Englewood Cliffs, New Jersey, 1972.

SACKMAN, H.
An Investigation of Certain Aspects of the Validity of the Formal Rorschach Scoring System in Relation to Age, Education, and Vocabulary Score
Unpublished doctoral dissertation, Fordham University, 1952.

SAHR, ROBERT C.
A Collation of Similar Delphi Forecasts
Institute for the Future, WP-5, April 1970.

A comparison of judgments made by different Delphi panels can have two main purposes. First, it can seek to illuminate differences of perspective among disciplines. Second, it can help identify differences that arise, not from panel assessments and judgments, but from extraneous factors. Judicious use of such comparisons can help to refine Delphi methodology so that its results are more reliable reflections of its goals--eliciting expert judgment.

Because the Delphi studies discussed here were not designed for comparison, this work, as the title states, is a very preliminary one. The author hopes that the approaches used here may suggest ways in which Delphi studies may be designed and conducted in the future so as to facilitate comparison.

The research presented in this paper was sponsored by the Educational Policy Research Center of the Syracuse University Research Corporation.

SALANCIK, G. R., T. J. GORDON, and N. ADAMS
On the Nature of Economic Losses Arising from Computer-Based Systems in the Next Fifteen Years
Institute for the Future, R-25, March 1972.

This study, sponsored by the Skandia Insce. Co. explores potential new uses of computers in the next fifteen years and the potential losses they could present in the event of misuse and malfunction. The report examines future growth of computer usage, risks attendant upon likely future computer applications and possible remedies for these risks. (B)

SALANCIK, G. R., W. WENGER, and E. HELFER
"The Construction of Delphi Statements"
Technological Forecasting and Social Change, Vol. 3, No. 1, 1971.

SCHMIDT, D. L.
Creativity in Industrial Engineering
The Rand Corporation, P-4601, March 1971.

A discussion of creativity and Delphi techniques to improve company research. The creative process synthesizes knowledge, logical reasoning, and originality--the mix depending on the field. In addition, the creative person uses multiple approaches, self-questioning, scepticism, nonconformity, and self-confidence and courage in presenting his ideas. Management techniques to motivate creativity are many, such as gearing the pressure to the goals, showing tolerance for failure, providing means of communication with colleagues, setting a creative atmosphere, welcoming disagreement, and accepting unconventionality. The Delphi procedure, by encouraging anonymity, controlled feedback, and statistical "group response" reduces the undesirable socially dominant and group pressure toward conformity. Delphi also includes some form of self-rating. In conjunction with the establishment of an environment and a management team, the Delphi technique could considerably improve the creative output of a company's research staff. (R)

SET INC.
See DELPHI PANEL ON THE FUTURE OF LEISURE AND RECREATION, SET INC.

SHAW, M.
Group Dynamics: The Psychology of Small Group Behavior
McGraw Hill, New York, 1971.

SHERIF, M.
The Psychology of Social Norms
Harper, New York, 1936.

SILLS, DAVID L. (Ed.)
International Encyclopedia of the Social Sciences
Macmillan, New York, 1968.

SKANDIA INSCE CO.
Sponsor of the Institute for the Future study R-25: *On the Nature of Economic Losses Arising from Computer-Based Services in the Next Fifteen Years*, (SALANCIK, GORDON, and ADAMS).

SMIL, V.
"Energy and the Environment--A Delphi Forecast"
Long-Range Planning, Vol. 5, No. 4, December 1972, pp. 27-32.

SMITH, KLINE, AND FRENCH LABORATORIES
See papers by BENDER and TEELING-SMITH. The Bender article gives detailed results from the study in graphical form. Also referenced in *Business Week* article, MARTINO, *Technological Forecasting*, and CETRON'S book. (B)

STEINER, GEORGE A.
Top Management Planning
The Macmillan Company, London, 1969.

STOGDILL, R. M.
Individual Behavior and Group Achievement
Oxford University Press, New York, 1959.

SULC, O.
Forecasting the Interactions Between Technological and Social Changes
Manchester Business School, University of Manchester, 061-237-4571, July 1968.

TEELING-SMITH, GEORGE
"Medicines in the 1990's: Experience with a Delphi Forecast"
Long-Range Planning, June 1971.
Describes a Delphi in the Office of Health Economics, London. References U.S. Delphi studies by Smith, Kline and French plus three unnamed other U.S. companies. (B)

"The Futures Business"
Chemical & Engineering News, Vol. 47, August 1969, pp. 62-75.
This article reviews the rapid growth of long-term forecasting activities in business, government, and policy research organizations. Recent (to 1969) books, journal articles, and studies are referenced. The examination of Delphi reviews the Rand and TRW experiences. One chart of TRW predictions is shown. (B)

THIESMEYER, LINCOLN R.
"How to Avoid Bandwagon Effect in Forecasting: The Delphi Conference Keeps Crystal Ball Clear"
Financial Post, October 1971.
A general review of the Delphi technique. The Institute for the Future Delphi studies are referenced. (B)

THIESMEYER, LINCOLN R.
"The Art of Forecasting the Future is Losing Its Amateur Status"
Financial Post, June 1971, p. 7.
Describes the growth of "futurology" and technological forecasting in business and the creation of the Institute for the Future with corporate help. References IFF Delphi on residential housing sponsored by Owens-Corning Fiberglas Corporation. Also mentions activities at General Electric, Goodyear Tire and Rubber, and Delphi by Smith, Kline and French: "A Delphi Study of the Future of Medicine." (B)

THOMPSON, L. T.
A Pilot Application of Delphi Techniques to the Drug Field: Some Experimental Findings
The Rand Corporation, R-1124, June 1973.
Report on an experiment designed to estimate the feasibility and usefulness of employing Delphi techniques to examine drug-related issues. The experiment consisted of two Delphi rounds and used a questionnaire which included both descriptive and evaluative items pertinent to drug use and drug programs. Drug-related questions are particularly prone to ambiguity and interpretive difficulties; a useful side benefit of Delphi in this context is the capacity to modify or augment questions on the basis of first-round responses. Although some convergence behavior was observed on almost all questions, the degree of convergence differed significantly among items.

THORNDIKE, ROBERT L.
Personnel Selection
Wiley, New York, 1949.

TRANS CANADA TELEPHONE SYSTEM
Communications, Computers, and Canada
December 1971.

This report by the TCTS (8 major Canadian Telephone Companies) outlines their commitment to long-range planning for computer and visual communications services. Three Bell Canada Business Planning Delphi studies are abstracted in Appendix B. Sample results are shown from the Education, Medicine, and Business studies (see DOYLE and GOODWILL, and GOODWILL references). (B)

TRW DELPHI STUDIES

See works by NORTH and/or PYKE. Results also shown in WILLS: *Technological Forecasting* and CETRON's book. Many of the other articles in this bibliography have brief references to the TRW Delphi work. Footnote No. 16 lists several other references not outlined in the bibliography. (B)

TUROFF, M.
"An Alternative Approach to Cross-Impact Analysis"
Technological Forecasting and Social Change, Vol. 3, No. 3, 1972, pp. 309-339.

Reviews the literature on the subject and the use of cross-impact in an information system context.

TUROFF, M.
"Delphi Conferencing (i.e., Computer Based Conferencing with Anonymity)"
Technological Forecasting and Social Change, Vol. 3, No. 2, 1972, pp. 159-204.

TUROFF, M.
The Design of a Policy Delphi
National Resource Analysis Center, Systems Evaluation Division, Executive Office of the President, Office of Emergency Preparedness, T.M. 123, 1970. (See also *Technological Forecasting and Social Change*, Vol. 2, No. 2, 1970, pp. 149-171.)

This paper is a comprehensive and definitive examination of the use of the Delphi method in policy issues. It develops the thesis that the Delphi could be useful as an adjunct to the committee approach in policy formulation; it is argued that the two are complimentary, rather than one being superio

The paper opens with a review of Delphi, argues for its use in policy decisions while pointing out possible dangers, and recommends some procedural guidelines.

There is a highly complete and up-to-date bibliography included. (P)

TUROFF, M.
"Delphi and its Potential Impact on Information Systems"

AFIPS Conference Proceedings, Vol. 39, pp. 317-326, AFIPS Press, Montvale, New Jersey, 1971.

VOYER, R. D.
"Inventing Future with Customade Technology"
Science Forum, Vol. 3, No. 5, 1970, p. 8.

VOYER, R. D.
"Delphi Technique--A Valuable Tool for Technological Forecasting"
Science Forum, Vol. 2, No. 5, 1969, p. 6.

WEAVER, W. T.
Delphi, A Critical Review
Syracuse University Research Corporation, RR-7, February 1972.

WEAVER, W. T.
Delphi as a Method for Studying the Future: Testing Some Underlying Assumptions
Educational Policy Research Center, Syracuse, New York, 1970.

WEAVER W. T.
An Exploration Into the Relationship Between Conceptual Level and Forecasting Future Events
Doctoral Dissertation, Syracuse University, New York, 1969.

WEYERHAEUSER COMPANY
This study is mentioned in the *Business Week* article. The Delphi research is into the future of the construction business. (B)

WHIRLPOOL
See the paper by DAVIS for Whirlpool's use of Delphi as a part of their information gathering system. (B)

WILCOX, W.
Forecasting Asian Strategic Environments for National Security Decisionmaking: A Report and a Method
The Rand Corporation, RM-6154, June 1970.

WILLS, GORDON
"The Preparation and Development of Technological Forecasts"
Long-Range Forecasting, March 1970.
A review of technological forecasting and planning in business. Several techniques are reviewed. Delphi results from a computer study by ICL are shown. The Hercules Powder Company study is referenced as well and results shown. The first round Hercules questionnaire is shown in an appendix. (B)

WILLS, GORDON, RICHARD WILSON, NEIL MANNING, and ROGER HILDEBRANDT
Technological Forecasting
Penguin Books, Middlesex, England, 1972.
This book subtitled "The Art and its Managerial Implications" examines technological forecasting from an essentially marketing viewpoint. The chapter on Delphi uses business examples. The Hercules

study is cited, showing the first round questionnaire (pp. 194-203) study design, and sample results (pp. 209-210). A sample feedback from a Ling-Temco-Vaught, Inc. (LTV) Study is also shown (p. 211). Three charts illustrate the ICL study (pp. 212-215). The chapter concludes with a discussion of the TRW experience, including charts (pp. 216-223). (B)

ZARNOWITZ, VICTOR
"On the Accuracy and Properties of Short-Term Economic Forecasts"
The Task of Economies, Forth-Fifth Annual Report of the National Bureau of Economic Research, 1965.